

**JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH  
EKONOMICKÁ FAKULTA**

**PREDIKCE FINANČNÍ TÍSNĚ PODNIKU  
Disertační práce**

**Ing. Radek Zdeněk**

**2012**

Studijní program: Ekonomika a management  
Studijní obor: Řízení a ekonomika podniku  
Školící pracoviště: Katedra účetnictví a financí  
Školitel: doc. Ing. Milan Jílek, Ph.D.

Prohlašuji, že svoji disertační práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

20. 4. 2012

Děkuji prof. Ing. Františku Střelečkovi, CSc., dr.h.c. a doc. Ing. Milanu Jílkovi, Ph.D. za odborný dohled, rady a připomínky poskytnuté během studia.

# Obsah

1.	Úvod	7
2.	Finanční tíseň a úpadek podniku	9
2.1.	Predikce finanční tísně a úpadku podniku	9
2.2.	Definice úpadku	9
2.3.	Definice finanční tísně	10
2.4.	Účetní výkazy jako zdroj dat	13
2.5.	Výběr proměnných	15
2.6.	Specifika odvětví zemědělství	17
3.	Klasifikační metody	18
3.1.	Problematika klasifikačních metod	18
3.2.	Tvorba klasifikačního modelu	18
3.3.	Profilová analýza	20
3.4.	Vícerozměrná diskriminační analýza	21
3.4.1.	Kanonická diskriminační analýza	21
3.4.2.	Klasifikační diskriminační analýza	25
3.4.3.	Porušení předpokladů	28
3.4.4.	Robustní MCD-odhady	30
3.4.5.	Jádrové odhady hustoty	31
3.4.6.	Metoda nejbližších sousedů	32
3.4.7.	Metoda nejbližších prototypů	33
3.5.	Logistická regrese	35
3.6.	Probitová regrese	37
3.7.	Neuronové sítě	37
3.7.1.	Matematický model neuronové sítě	38
3.7.2.	Algoritmus učení Back-propagation	41
3.7.3.	Genetické učení	42
3.8.	Klasifikační stromy a klasifikační lesy	43
3.8.1.	Klasifikační lesy	45
3.9.	Hodnocení kvality klasifikátorů	47
3.9.1.	Vyhodnocení účinnosti klasifikace	47
3.9.2.	Klasifikační matice	48
3.9.3.	ROC křivky	49
3.9.4.	Skutečný podíl chyb	52
4.	Klasifikační modely	53
5.	Metodika a materiál	72
5.1.	Vymezení podniků ohrožených finanční tísní	72
5.2.	Výběr ukazatelů	73
5.3.	Dílčí cíle	75
5.4.	Materiál	76
5.5.	Softwarové vybavení	77
6.	Řešení a výsledky	78
6.1.	Využitelnost stávajících klasifikačních modelů	78
6.2.	Profilová analýza	79
6.3.	Kombinace jednorozměrných klasifikačních pravidel	81
6.3.1.	Kombinace třech ukazatelů	83

6.3.2.	Kombinace pěti ukazatelů	83
6.3.3.	Kombinace sedmi ukazatelů	84
6.3.4.	Kombinace devíti ukazatelů	85
6.4.	Klasifikační síla ukazatelů	85
6.5.	Výběr ukazatelů pro vícerozměrné klasifikační metody	87
6.5.1.	Grubbsův test	89
6.5.2.	Test normality	89
6.5.3.	Testy shody středních hodnot	90
6.6.	Diskriminační analýza	91
6.6.1.	Lineární diskriminační analýza	91
6.6.2.	Kvadratická diskriminační analýza	94
6.6.3.	Robustní diskriminační analýza	95
6.7.	Logistická regrese	97
6.8.	Probitová regrese	101
6.9.	Metoda nejbližších sousedů	103
6.9.1.	Metoda nejbližšího souseda	103
6.9.2.	Metoda $k$ -nejbližších sousedů	105
6.10.	Metoda nejbližších prototypů	108
6.11.	Neuronové sítě	109
6.11.1.	Neuronové sítě s 96 vstupními neurony	110
6.11.2.	Neuronové sítě s 11 vstupními neurony	110
6.11.3.	Neuronové sítě s 6 vstupními neurony	112
6.11.4.	Neuronové sítě s 3 vstupními neurony	114
6.12.	Klasifikační stromy a klasifikační lesy	117
6.12.1.	Klasifikační stromy	117
6.12.2.	Klasifikační les	119
7.	Závěr	122
	Souhrn	126
	Summary	127
	Literatura	128
	Seznam obrázků	137
	Seznam tabulek	138

# 1. Úvod

Téma predikce úpadku podniku se v ekonomické literatuře objevuje od šedesátých let minulého století a byla mu věnována řada vědeckých prací. S cílem předpovědět úpadek podniku nebo správně klasifikovat podnik podle jeho finančního zdraví byla využívána řada různých metod, které se vzájemně liší jak svými předpoklady, tak složitostí.

Predikční modely jsou založeny na hypotéze, že finanční obtíže podniku lze podle určitých signálů, které se mohou projevit v hodnotách poměrových finančních ukazatelů, identifikovat dříve, než se skutečně projeví. Cílem zkoumání je pak vyvinout takový model, který dokáže vydávat varovné signály s dostatečným časovým předstihem - model včasného varování.

Již v roce 1966 William Beaver hledal takový poměrový ukazatel, který by lépe než jiné dokázal s dostatečným předstihem upozornit na blížící se úpadek podniku. Z poměrových ukazatelů pak odvodil několik jednorozměrných diskriminačních modelů. O dva roky později Edward Altman využil pro řešení obdobného problému vícerozměrnou diskriminační analýzu a na jejím základě odvodil známý Z-score model. Altmanův model se stal etalonem oboru a v řadě vědeckých prací je využíván jako základ pro komparaci výsledků. Profesor Altman se tématu věnuje do současnosti, přičemž dosud představil několik modifikací původního modelu. Z hlediska metodického aparátu diskriminační analýza nad ostatními klasifikačními metodami převládá, rozšiřuje se využití dalších metod, jako je logistická nebo probitová regrese. Kromě těchto metod je možné využít další klasifikační metody, zejména neuronové sítě či klasifikační stromy. Do současnosti byla publikována řada komparačních studií, ve kterých byla srovnávána spolehlivost klasifikace jednotlivých metod a modelů. Jejich závěry jsou však často protichůdné a na jejich základě je obtížné obecně doporučit určitou klasifikační metodu či model.

Význam predikce finanční tísně nebo úpadku podniku v České republice dokumentuje tisková zpráva Czech Credit Bureau (CCB 2011). Od roku 1993 do poloviny roku 2011 zbankrotovalo v Česku 2 674 akciových společností a 16 105 společností s ručením omezeným. V další zprávě (CCB 2012) uvádějí nejnovější údaje pro rok 2011. V roce 2011 bylo vyhlášeno 2 413 firemních bankrotů (meziroční nárůst o 49 %), z toho 1 232 se týkalo právnických osob a 1 181 fyzických osob podnikatelů. Z odvětvového hlediska bylo v roce 2011 vyhlášeno nejvíce bankrotů na firmy v odvětví obchod a pohostinství (416), u subjektů podnikajících ve službách to pak bylo 367 bankrotů. K nejvyššímu meziročnímu růstu v počtu bankrotů došlo v odvětvích zemědělství a těžba (nárůst o 40 %); absolutní čísla však nejsou uvedeny.

Tisková zpráva České kapitálové informační agentury (Čekia 2010) konstatuje, že zemědělství, zpracovatelský průmysl, obchod a doprava patří v České republice k nejrizikovějším oborům podnikání, naopak nejstabilnější jsou podniky v odvětvích finanční služby, školství, zdravotnictví a energetika. Výsledky vycházejí z metodiky ratingového modelu ČEKIA Stability Rating, pomocí kterého

je odhadováno riziko úpadku podniku v následujícím roce. V odvětvích zemědělství, myslivost, lesnictví, rybolov a chov ryb je 14,3 % podniků zařazeno mezi vynikající (A), 51,9 % mezi dobré (B) a 32,6 % mezi rizikové a v úpadku (C a D; zbytek 1,2 % je neklasifikován). Nejnovější tisková zpráva (Čekia 2012) doplňuje, že podnikání v odvětví zemědělství patří stále k nejrizikovějšímu. Výsledky hodnocení podle Čekia Stability Rating jsou uvedeny v tabulce (Tab. 1). Oproti roku 2010 se podíl rizikových podniků v odvětví zemědělství zvýšil o 16 procentních bodů a průměrnou hodnotu v České republice převyšuje 2,2krát.

Tab. 1 Rozdělení podniků dle ČEKIA Stability Rating

Ratingové hodnocení	Odvětví zemědělství	Podnikatelská základna ČR
A - vynikající	13,74 %	21,67 %
B - dobré	36,70 %	55,59 %
C - rizikové	48,82 %	21,93 %
D - úpadky	0,74 %	0,82 %

Pramen: Čekia (2012)

Čekia (2012) nastiňuje hlavní příčiny těchto nepříznivých stavů. Lze je identifikovat především v silících konkurenčních tlacích a dodatečných investicích nutných pro dodržení platných norem kladených na welfare zvířat a ochranu životního prostředí. Po vstupu do Evropské unie a otevření trhu je české zemědělství konfrontováno s nabídkou zahraničních potravin, nedostatečný odbyt na domácím trhu a nestabilní výkupní ceny řeší některé podniky útlumem nebo ukončením produkce nerentabilních komodit, případně ukončením činnosti. Pokles objemu produkce zemědělských komodit se týká zejména živočišné výroby - především chovu prasat a chovu skotu. Podíl zemědělství na celkové produkci se snížil z 3,11 % v roce 2000 na 1,97 % v roce 2009.

**Cílem disertační práce** je ověření možností vícerozměrných klasifikačních metod při predikci finanční tísně podniků v odvětví zemědělství.

Samotná práce bude strukturována do několika základních částí. Literární přehled je možné rozdělit na tři části. V první bude shrnut dosavadní stav poznání v oblasti finanční tísně a její predikce; druhá bude obsahovat popis vybraných klasifikačních metod; a třetí popis vybraných klasifikačních modelů sestavených pro předpověď úpadku a finanční tísně nebo k hodnocení finančního zdraví podniku. V metodické části budou nadefinovány proměnné (závislá i nezávislé) a dále bude popsán materiál a postup řešení. V aplikační části bude nejprve ověřena využitelnost stávajících klasifikačních modelů pro predikci finanční tísně. Následně budou konstruovány predikční modely s využitím aparátu klasifikačních metod popsaných v literárním přehledu.



## 2. Finanční tíseň a úpadek podniku

### 2.1. Predikce finanční tísně a úpadku podniku

Modely předpovědi finanční tísně i úpadku podniku jsou založeny na klasifikační proceduře, na jejímž základě je objekt (tedy podnik) zařazen do jedné z předem známých skupin (například ohrožený úpadkem nebo prosperující). Mohou nastat čtyři situace, tedy že

1. ohrožený podnik je správně klasifikován mezi ohrožené podniky;
2. ohrožený podnik je mylně klasifikován mezi prosperující podniky;
3. prosperující podnik je chybně zařazen do skupiny ohrožených podniků;
4. prosperující podnik je klasifikován korektně.

Tyto situace lze přehledně zobrazit v klasifikační (též konfusní) matici (v tomto případě v matici typu  $2 \times 2$ ). Tvorba a hodnocení klasifikační matice jsou blíže popsány v části 3.9.2 Klasifikační matice). Pokud je ohrožený podnik klasifikován jako prosperující, pak nastává chyba I. druhu; pokud je prosperující podnik mylně zařazen do skupiny ohrožených, pak nastává chyba II. druhu. S chybou I. i II. druhu mohou být spojeny určité náklady, které je možné (dokonce vhodné) v některých klasifikačních metodách zohlednit.

### 2.2. Definice úpadku

Rýdl (2005) uvádí, že kvalitní úpadkové právo umožňuje rychlý výstup problematických subjektů z hospodářského systému, a tím omezuje dopady vzniklé činností úpadců na ostatní podnikatele, domácnosti a finanční instituce.

Zákon č. 182/2006 Sb., o úpadku a způsobech jeho řešení (insolvenční zákon) definuje v §3 odst. 1, že dlužník je v úpadku, pokud má:

- a) více věřitelů a
- b) peněžité závazky po dobu delší 30 dnů po lhůtě splatnosti a
- c) tyto závazky není schopen plnit.

V odst. 2 upřesňuje, že dlužník není schopen plnit své peněžité závazky, pokud:

- a) zastavil platby podstatné části svých peněžitých závazků, nebo
- b) je neplní po dobu delší 3 měsíců po lhůtě splatnosti, nebo
- c) není možné dosáhnout uspokojení některé ze splatných peněžitých pohledávek vůči dlužníku výkonem rozhodnutí nebo exekucí, nebo
- d) nesplnil další povinnosti dané zákonem.

Odst. 3 rozšiřuje definici úpadku o předlužení. O předlužení jde podle zákona tehdy, má-li dlužník více věřitelů a souhrn jeho závazků převyšuje hodnotu jeho majetku.

Insolvenčním zákonem byl nahrazen zákon č. 328/1991 Sb., o konkurzu a vyrovnání. Tomuto zákonu bylo vytýkáno, že vycházel z práva a etiky první republiky a hodil se k realizaci bankrotu spíše malého podniku s maximálně desítkami věřitelů. Zákon neumožňoval alternativy k likvidačnímu způsobu řešení úpadku, nerozlišoval mezi typy úpadců a neposkytoval věřitelům dostatečné právní mechanismy při rozhodování o způsobu řešení úpadku (Rýdl 2005).

### 2.3. Definice finanční tísně

Kritérium finanční tísně bývá v literatuře definované rozličným způsobem, často je založeno na ztrátách trvajících určitou dobu (ať provozní či celkové); nevyplacení dividend z prioritních akcií; neplnění závazků z dluhopisů; rozsáhlém propouštění zaměstnanců; kapitálové restrukturalizaci; kumulovaných ztrátách či negativním cash flow (Balcaen a Ooghe 2006).

Abou El Sood (2008) radí společnost mezi společnosti ohrožené finanční tísní, pokud splňuje během tří po sobě následujících roků jednu z podmínek: 1) záporné provozní cash flow v kterémkoliv roce; 2) čistá provozní ztráta v kterémkoliv roce; nebo 3) záporný čistý pracovní kapitál v kterémkoliv roce.

Podobně vymezuje problémové zemědělské podniky Kopta (2009), které jsou ohroženy jak dlouhodobou zápornou rentabilitou (součet hospodářských výsledků za pět let je záporný), tak prudkými výkyvy hospodářského výsledku, které vedou k zápornému provoznímu cash flow.

Gurčík (2002) za prosperující podniky považoval ty, které v letech 1998 až 2000 dosahovaly zisku a v posledním roce rentabilita vlastního kapitálu byla vyšší než 8 % (míra inflace v roce 2000), za neprospěrující ostatní podniky.

Synek (1996) uvádí tři situace, ve kterých lze podnik označit jako ohrožený finanční tísní.

- Podnik je dlouhodobě ztrátový a platebně neschopný.
- Podnik se vyznačuje dílčími vážnými problémy, zejména poklesem objemu výkonů, vznikem nepříznivého sociálního klimatu ve společnosti, trvalou platební neschopností.
- Podnik se jeví jako úspěšný, ale v důsledku extrémního růstu trvale naráží na nedostatek kapitálu.

Příznaky nenastávají současně, ale v určitých fázích. Nejprve jde o pokles objemu výkonů, dále nastupuje pokles rentability, zvýšená potřeba pracovního kapitálu, zhoršování kapitálové struktury a nakonec dochází k trvalé platební neschopnosti.

Blum (1974) charakterizuje podnik jako určitý zásobník finančních prostředků, přičemž pravděpodobnost finanční tísně závisí na jejich očekávaných tocích. Pravděpodobnost finanční tísně je pak tím vyšší, čím

- a) menší je tento zásobník;
- b) menší jsou provozní příjmy;
- c) větší jsou závazky k věřitelům;
- d) větší jsou provozní výdaje;
- e) vyšší je variabilita příjmů a výdajů;
- f) „náchylnější k úpadku“ je odvětví, ve kterém firma podniká. V těchto odvětvích je četnost úpadku vyšší než v ostatních odvětvích. Jako příklad autor uvádí výrobu automobilů na počátku dvacátého století.

Marek (2006) dělí finanční tíseň do dvou podob:

- Relativní, která nastává v okamžiku, kdy podnik není schopen dostát k danému termínu všem svým splatným závazkům.
- Absolutní, tj. stav, kdy hodnota závazků podniku převyšuje hodnotu jeho aktiv.

Podle Bakera a Powella (2005) finanční tíseň nastává, pokud má společnost obtíže s plněním smluvních závazků spojených s dluhovým financováním. Jedná se o celkové oslabení finanční kondice podniku způsobené enormním zadlužením, přičemž krajním případem finanční tísně je bankrot.

Zřídka kdy se stane, že jedno chybné rozhodnutí zavede společnost do potíží, obvykle je to celá série mylných rozhodnutí, popř. rozhodovací procesů, které se neprojeví okamžitě (Jindřichovská a Blaha 2001).

Nevýhodou finančních predikčních modelů může být zanedbání nefinančních informací o podniku. Odborníci se domnívají, že problémy podniků jsou spojeny s chybami v jejich řízení a tyto chyby se objeví dříve, než se projeví v číslech, tedy ve finančních výkazech. Jedním z nefinančních predikčních modelů je A skóre. Profesor Argenti<sup>1</sup> v tomto modelu stanovil určité symptomy, které považoval za klíčové pro finanční situaci podniku a každému přiřadil určitou váhu (Sůvová 2000).

Lussier a Corman (1994) uvádějí patnáct příčin úpadku, přičemž řadu z nich lze uvažovat pouze v případě podniku fyzických osob (kapitál, finanční kontroly, zkušenosti v odvětví, manažerské zkušenosti, plánování, odborní poradci, vzdělání, zaměstnanci, načasování produktu, ekonomický cyklus, věk, partneři, rodiče, menšiny a marketing).

Valach (1999) člení příčiny finanční tísně na interní (vyvolané přijetím špatného rozhodnutí v rámci podniku) a externí (dané faktory existujícími nezávisle na jednání osob spojených s podnikem). Dále uvádí náměty na řešení finanční tísně (vypracování nové obchodní strategie; zlepšení hospodaření s pohledávkami, se zásobami; zvýšení příjmů prodejem aktiv; snížení či odložení výdajů; zaměření se na rychlý přísun peněz pomocí skont; sestavení žebříčku priorit ve vlastních platbách; vypracování kritérií o zastavení či omezení provozu podniku).

---

<sup>1</sup> Původní pramen ARGENTI, J. Corporate collapse: the causes and symptoms. London: McGraw-Hill, 1976.

Giroux (2003) mezi příznaky finanční tísně řadí nízkou likviditu, vysoké zadlužení, nízké či kolísavé zisky, klesající tržby a záporné provozní cash flow.

Grünwald (2001) definuje finanční tíseň jako podlomení finančního zdraví, přičemž finanční zdraví je dáno aktuálním stavem podnikových financí. Finanční tíseň je opačným extrémem finančního stavu podniku, jakým je naprosté finanční zdraví. Mezi těmito extrémy existuje nekonečně mnoho stavů, které lze označovat různými názvy (Grünwald a Holečková 2004). Finančně zdravý podnik nejeví příznaky finančního ohrožení svého pokračujícího trvání. Lze předpokládat, že v dohledné době (minimálně do roku) nedojde k platební neschopnosti ani k předlužení. Podnik je ve finanční tísní, když vykazuje tak vážné platební potíže, které nemohou být vyřešeny bez radikálních změn v provozní nebo ve finanční činnosti (Holečková 2008). Míra finančního zdraví podniku má vyjádřit, jak velkým rizikům z provozní činnosti pravděpodobně finance podniku odolají. Je vyjadřována zařazením podniku do jednoho ze čtyř pásem:

- Pevné zdraví. Pevné zdraví by jistilo podnik i při závažných nezdarech v provozní činnosti či při externím ohrožení.
- Dobré zdraví. Dobré zdraví by podrželo podnik při přechodných nesnázích v provozní činnosti. Čím je blíže k úrovni pevného zdraví, tím větší je odolnost pro externímu ohrožení.
- Slabší zdraví. Při slabším zdraví by případné poruchy v podnikání mohly způsobit přechodné finanční potíže.
- Churavění. Churavění přivádí podnik do finanční tísně. Nelze vyloučit, že dojde k úpadku.

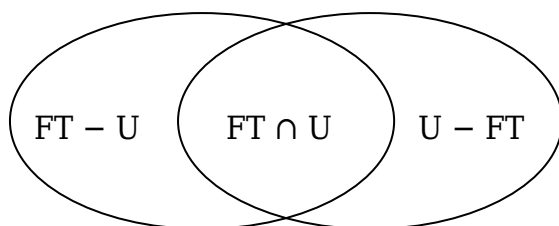
Z uvedeného přehledu vyplývá, že definice jevu finanční tísně není jednotná a je stále předmětem zkoumání. Pojmy finanční tíseň a úpadek podniku se v literatuře často nerozlišují. Dále je možné setkat se s termínem default, který je definován jako nedodržení závazku dlužníka vyplývajícího z úvěrové smlouvy (Jakubík a Teplý 2007). I v anglofonní literatuře se v souvislosti s úpadkem, finanční tísní a finančním zdravím podniku objevují termíny, jejichž významy se částečně překrývají. Jejich významy uvádí Velký ekonomický slovník (Bürger 2007):

- bankruptcy - bankrot, konkurz (soudní řízení), úpadek (finanční);
- failure - krach, neúspěch, nezdar, selhání, ztroskotání (projektu);
- distress - nouze, zabavení, obstavení, tíseň;
- default - insolvence, nedodržení lhůty, nedodržení závazku, neplacení, neplnění, prodlení v placení, nedodržet lhůtu, nedodržet závazek;
- insolvency - nesolventnost, platební neschopnost, úpadek;
- financial health, financial sound - finanční stabilita, finanční zdraví.

Vyhlášení úpadku (bankrotu, konkurzu, insolvence) je obvykle spojeno s nedostatečnou likviditou a insolvenčí, analyzovaný výběrový soubor podniků však může být kontaminován podniky, které jsou sice formálně bankrotní, ovšem nejeví příznaky finanční tísně (Balcaen a Ooghe 2006). Téměř všechny bankrotní podniky se nacházejí ve finanční tísní, ovšem ne na všechny podniky ve finanční tísní je vyhlášen bankrot (Grice a Dugan 2001). Tuto situaci lze zob-

razit jako průnik množin podniků ve finanční tísně (FT) a podniků v úpadku (U, Obrázek 1).

Obrázek 1 Průnik finanční tísně a úpadku



Pramen: vlastní

Úpadek podniku může nastat i vlivem neočekávané události, např. živelnou pohromou či nepříznivými klimatickými podmínkami. Zařazení takovýchto podniků do souboru upadajících zvyšuje šum a snižuje predikční schopnost odvozeného modelu. Vyhlášení úpadku je jednou z několika možností řešení finanční tísně podniku - v úvahu dále přichází sloučení s jiným podnikem, rozdělení, likvidace. Další zkruslení vyplývá z možné časové prodlevy mezi vznikem finančních obtíží a vyhlášením bankrotu (Balcaen a Ooghe 2006).

Provedení provozní a finanční restrukturalizace a reorganizace by mělo zastavit a odvrátit směřování k úpadku či zániku. Zánikem podniku je ohrožen management, zaměstnanci i externí partneři, kteří spoléhají na finanční analýzu jako na nástroj včasného varování. Výstraha by měla přijít s takovým předstihem, aby se management mohl pokusit úpadek odvrátit a aby se vlastníci, věřitelé a obchodní partneři mohli chránit před nepříznivými následky (Holečková 2008).

## 2.4. Účetní výkazy jako zdroj dat

Podnik a jeho fungování lze popsat pomocí řady finančních i nefinančních charakteristik. Nejkoncentrovanějším obrazem fungování podniku je obraz v podobě dat ze základních účetních výkazů - rozvahy a výkazu zisku a ztráty (Neumaierová a Neumaier 2008). Účetní výkazy představují základní zdroj dat pro výpočet poměrových ukazatelů, mají však některé slabé stránky, které jsou dány jejich konstrukcí a účetními praktikami. Účetní výkazy pracují s údaji, které se vztahují k minulosti. Analýza vždy čerpá z více či méně vzdálené minulosti a pracuje s rizikem, že stav zachycený ve výkazech a aplikovaný na dnešní hodnocení nemusí být aktuální (Blaha a Jindřichovská 2006).

Rozvaha charakterizuje situaci podniku k určitému okamžiku, podává informaci o celkových aktivech a pasivech a jejich struktuře. S její konstrukcí jsou spojené slabé stránky (Růčková 2010, Blaha a Jindřichovská 2006):

- Účetní standardy používají jako základ pro ohodnocení rozvahových položek historickou hodnotu – dřívější pořizovací cenu. Ačkoliv je tato cena upravována odpisy, nereflktuje přesně skutečnou současnou hodnotu.
- K určení hodnoty některých rozvahových položek musí být užit odhad (např. rozdíl účetní hodnotou a aktuální hodnotou).
- Všeobecnou praxí je postupné snižování hodnoty dlouhodobých aktiv – jejich odepisování. Opačný proces, tedy jejich zhodnocování, se nebere v úvahu.
- V rozvaze nejsou zahrnuty některé položky, ačkoliv mají určitou finanční hodnotu a účastní se hospodářského procesu, např. majetek pořízený pomocí finančního leasingu.

Snahou výkazu zisku a ztráty je změřit čistý zisk, jakožto výsledek za určité časové období. Je sestaven na tokovém základě, na rozdíl od rozvahy. Jde o zachycení souvislosti mezi výnosy podniku dosaženými v určitém období a náklady spojených s jejich tvorbou. Změny v čase nemusejí být rovnoměrné – výnosy dosažené v určitém období a náklady s nimi spojené nemusejí být vynaloženy ve stejném období (existují účty časového rozlišení výnosů a nákladů). Nákladové a výnosové položky se neopírají o skutečné peněžní toky, výsledný čistý zisk tedy neobsahuje skutečné peněžní prostředky získané hospodařením v daném období (Růčková 2010).

Přehled o peněžních tocích (výkaz cash flow) srovnává bilanční formou zdroje tvorby peněžních prostředků (příjmy) s jejich užitím (výdaji) za určité období. Přehled o peněžních tocích poskytuje informace o struktuře finančních zdrojů získaných podnikem a o struktuře jejich užití v daném období. Výhodou výkazu cash flow je, že není ovlivněn metodou odpisování majetku a není zkreslován systémem a větší časového rozlišení nákladů a výnosů (Růčková 2010).

Svou roli hraje také věrohodnost finančního výkaznictví a vypovídací schopnost primárních ukazatelů (a tedy i ukazatelů sekundárních), se kterými klasifikační metody a modely pracují (Jindřichovská a Blaha 2001). Zajištění vypovídací schopnosti účetních výkazů je základní podmínkou komunikace mezi jejich uživateli. Vypovídací schopnost údajů však snižuje řadu skutečností. V praxi tak vznikají základní problémy – účetní výkazy nedokumentují přesně ekonomickou realitu hospodaření podniků; druhý problém vyplývá z nejednotnosti pravidel a výkaznictví v různých zemích (Knápková a Pavelková 2010). Drábková a Kouřilová (2008) navíc přidávají úmysl prostřednictvím povolených variant v účetnictví reagovat na požadovaný cíl (např. vytvořit příznivější obraz o finanční pozici podniku v účetních výkazech oproti skutečnosti), který může přecházet až na úroveň podvodů.

Rozvoj globálního prostředí a působení firem vyvolává potřebu nadnárodních, celosvětově platných a uznávaných účetních norem. Situace v této oblasti je složitá – ačkoliv v 70. letech minulého století se začaly objevovat snahy o mezinárodní harmonizaci účetnictví a v dalších letech došlo k výraznému pokroku, nejsou zatím účetní normy celosvětově platné a závazné. Nejstarší systém představují národní účetní standardy USA – US GAAP. Tento systém výkaznictví

se snaží vyhovět požadavkům světových burz. V Evropské unii je pro podniky, které jsou emitenty cenných papírů registrovaných na burzách cenných papírů v členských státech EU, povinné použít pro účtování a sestavení účetní závěrky Mezinárodní účetní standardy (IAS/IFRS). Současně musí české podniky pro daňové účely sestavovat účetní závěrku i podle české účetní legislativy (Knápková a Pavelková 2010).

## 2.5. Výběr proměnných

Při tvorbě predikčního modelu je obvyklý postup, kdy se vychází z počáteční řady ukazatelů, které jsou vybrány např. podle jejich dosavadního použití v předchozích studiích. Z této počáteční množiny ukazatelů se na základě statistických metod (testy diskriminační síly individuálních ukazatelů, krokové metody výběru, výběr nejlepší podmnožiny, korelační analýza aj.) provádí výběr konečné (Balcaen a Ooghe 2006; Štreleček a Zdeněk 2004). Takovýto výběr ukazatelů je závislý na výběrovém souboru a výsledky lze těžko zobecňovat či aplikovat na jiný soubor (Grice a Ingram 2001).

Vysoká multikolinearita ukazatelů zařazených do modelu pak např. u diskriminační analýzy způsobuje, že koeficienty diskriminátorů mají neintuitivní znaménka. Obecně převažuje názor, že v predikčním modelu by měly být zastoupeny ukazatele z celého spektra finanční analýzy (Balcaen a Ooghe 2006).

McLeay a Omar (2000) rozdělují poměrové ukazatele na ohraničené (bounded) a neohraničené (unbounded). Ohraničené ukazatele jsou zdola omezené nulou, extrémních hodnot mohou nabývat pouze na pravém chvostu; neohraničené nemají horní ani spodní mez a extrémních hodnot mohou nabývat na obou stranách rozdělení. Pro splnění podmínky některých klasifikačních metod (normality) autoři navrhují pro ohraničené ukazatele použití Box-Coxovy mocninné transformace, pro neohraničené ukazatele transformace snižující špičatost. Užití takto transformovaných proměnných zvyšuje spolehlivost klasifikace u modelů vytvořených jak pomocí diskriminační analýzy, tak i logistické regrese.

Štreleček a Zdeněk (2005, 2006) člení základní finanční poměrové ukazatele do skupin podle úrovně jejich standardizace, a to ukazatele standardizované na okolí hodnoty, na rozpětí a nestandardizované. Jako příklad lze uvést ukazatele zadluženosti. Předpokládejme, že  $VK \geq 0$  a  $VK + CK = K$ , pak oborem hodnot ukazatele  $CK / K$  je interval  $\langle 0, 1 \rangle$ , zatímco u jeho reciproké hodnoty  $K / CK$  je oborem hodnot interval  $\langle 1, \infty \rangle$  a u ukazatele  $VK / CK$   $\langle 0, \infty \rangle$ <sup>2</sup>. Výběrem vhodné formy ukazatele pak lze předejít situaci, kdy extrémní hodnota nestandardizovaného dílčího ukazatele významným způsobem ovlivní hodnotu syntetického ukazatele, a tedy rozhodne o výsledku klasifikace.

---

<sup>2</sup>  $VK$  je vlastní kapitál,  $CK$  cizí kapitál a  $K$  celkový kapitál.

Dambolena a Khoury (1980) mezi vysvětlující proměnné zahrnují i ukazatele variability (směrodatnou odchylku, směrodatnou chybu odhadu lineárních trendů a variační koeficient), Abidali a Harris (1995) trendy poměrových ukazatelů.

Platt a Platt (1991) porovnávají modely, které obsahují pouze poměrové ukazatele s modely, ve kterých jsou hodnoty poměrových ukazatelů vztaženy k odvětvovým průměrům. Od takto upravených modelů očekávají jejich vyšší stabilitu, vyšší stabilitu regresních koeficientů v čase a menší rozdíly mezi ex ante a ex post klasifikací. Na jejich empirickém materiálu je při použití ukazatelů vztažených k odvětvovým průměrům v případě ex ante klasifikace spolehlivost o 4 procentní body vyšší.

Edmister (1972) hledá optimální diskriminátor pro úpadek malých podniků, a to a) z poměrových ukazatelů, b) z tříletých trendů poměrových ukazatelů, c) z tříletých průměrů poměrových ukazatelů a z a) - c) vztažených k odvětvovým průměrům.

Zařazení ukazatelů založených na hrubé přidané hodnotě zvyšuje spolehlivost klasifikačních modelů. Hodnoty těchto ukazatelů se však mezi jednotlivými odvětvími značně liší, což je vhodné zohlednit (Declerc et al. 1992).

Abou El Sood (2008) testuje, zda doplnění modelu založeného na účetních údajích o makroekonomické a strategické podnikové ukazatele zvyšuje přesnost klasifikace. Přesnost klasifikace u rozšířených modelů se v jeho případě významně nezlepšila. Podobně Agarwal a Taffler (2008) konstruují modely založené na tržních veličinách a provádějí srovnání s klasickým modelem založeným na účetních údajích. Přesnost klasifikace měřená pomocí AUC (více v části 3.9.3 ROC křivky, str. 49) je u klasického modelu 0,89, u tržních modelů maximálně 0,87.

Xu a Wang (2009) doplňují model založený na poměrových ukazatelích o technickou efektivnost (model s konstantními výnosy z rozsahu), vypočtenou na základě metody datových obalů. Becchetti a Sierra (2003) kromě poměrových finančních ukazatelů přidávají do modelu ukazatele jako koncentrace zákazníků či přítomnost konkurentů a také technickou efektivnost.

Peel a Peel (1987) doplňují ukazatele, které významně přispívají k predikci úpadku podniku, o dobu mezi koncem účetního roku a zveřejněním výroční zprávy.

Mossman et al. (1998) srovnávají modely založené na (1) poměrových ukazatelích (zvolených podle Altman (1968)), na (2) cash flow (z provozní, z investiční činnosti a další vztažené k účetní hodnotě aktiv), (3) výnosech akcií (za 12 a 60 měsíců) a (4) variabilitě výnosů akcií (za 12 a 60 měsíců). Modely (3) a (4) vykazují ve srovnání s modely (1) a (2) nižší spolehlivost. Podle Charitou et al. (2004) ukazatele založené na cash flow v některých odvětvích výrazně přispívají k diskriminační síle modelů predikce úpadku.

Blum (1974) navrhuje zařazení např. podílů či násobků tradičních ukazatelů finanční analýzy nebo primárních ukazatelů (zde však doporučuje pracovat s vyváženými soubory, kde jsou pozorování párována např. podle velikosti).



Jakubík a Teplý (2007) pro modelování pravděpodobnosti úpadku podniku nevyužívají hodnoty poměrových finančních ukazatelů, nýbrž hodnoty relativního pořadí v použitém datovém vzorku. Každá hodnota ukazatele byla převedena na číslo z intervalu 0 až 1, čímž je zajištěna vyšší robustnost modelu vůči odlehkým hodnotám ukazatele.

## 2.6. Specifika odvětví zemědělství

Vzhledem k faktu, že práce se zabývá predikcí finanční tísně zemědělských podniků, je vhodné zmínit i základní specifika tohoto odvětví. Agrární sektor, do kterého patří i zemědělská prvovýroba, plní dvě základní funkce. 1) Je výrobcem specifického zboží (potravin), které jsou nezastupitelné - uspokojují základní lidské potřeby. 2) Je rozhodujícím činitelem při tvorbě venkovského životního prostředí - plní úkoly při tvorbě krajiny, udržení kvality vod, čistoty vzduchu, zabezpečuje osídlení (Bečvářová 2001).

Zemědělská výroba je determinována biologickým reprodukčním cyklem. Délka výrobního cyklu je dána právě délkou biologického cyklu (délkou vegetačního období, délkou odchovu a chovu) s minimální možností zkrácení. Rostlinná výroba má sezónní charakter, kdy se výrobní cyklus nekryje s pracovním procesem. Zemědělská výroba se uskutečňuje v bezprostředním spojení s přírodou a vzájemně se ovlivňují. Přírodní podmínky určují charakter a strukturu výroby, typ zemědělské výroby pak ovlivňuje kvalitu přírody a životního prostředí (Bečvářová 2001).

Na zemědělskou výrobu působí přírodní podmínky, které je možné rozdělit na proměnlivé a konstantní. Mezi relativně konstantní patří zejména podnebí, hydrologické poměry, nadmořská výška, orientace pozemků a půdní typ. K proměnlivým faktorům se řadí především počasí a množství a rozložení srážek, teplot a slunečního svitu (Bečvářová 2001). Vliv klimatických podmínek vytváří specifické pracovní prostředí jak pro lidi, tak pro stroje, a zvyšuje rizikovitost výroby. Podniky v lepších půdních a klimatických podmínkách mohou dosáhnout vyšší produktivity práce a vyššího zhodnocení kapitálu (Rosochatecká et al. 1999).

## 3. Klasifikační metody

### 3.1. Problematika klasifikačních metod

V tradiční regresní analýze je cílem určit funkční vztah mezi závisle proměnnou a vektorem nezávisle proměnných na základě vzorku existujících pozorování. Většina klasifikačních metod je založena na podobném přístupu. Významným rozdílem mezi regresí a klasifikací je charakter vysvětlované proměnné (Doupoupos a Zopounidis 2002). V terminologii těchto metod se rozlišuje učení s učitelem a učení bez učitele (Řezanková et al. 2009).

#### Učení s učitelem

V případě učení s učitelem (supervised learning) obsahuje vstupní datový soubor informace o příslušnosti objektů do známých<sup>3</sup> skupin. Cílem je vytvořit model, na jehož základě by mohly být objekty bez známé příslušnosti zařazovány do daných skupin. Termín klasifikace se obvykle používá v užším slova smyslu pro metody vycházející z principu učení s učitelem (Řezanková et al. 2009).

#### Učení bez učitele

Při učení bez učitele (unsupervised learning) není předem známa příslušnost žádného z objektů a obvykle není znám ani počet skupin (Řezanková et al. 2009). Učení bez učitele zahrnuje jednak shlukování či segmentaci (objektů, proměnných i kategorií), jednak redukci dat (proměnných či kategorií) (Řezanková a Húsek 2001).

### 3.2. Tvorba klasifikačního modelu

Schéma (Obrázek 2) znázorňuje postup při tvorbě klasifikačního modelu, který lze rozdělit na dvě části, a to na trénování a testování. Další fází je aplikace vytvořeného modelu. Trénováním se rozumí tvorba klasifikačního modelu na základě trénovacího souboru. Testováním se ověřuje kvalita vytvořeného klasifikačního modelu. Ve schématu (Obrázek 2) je závisle proměnná, která určuje příslušnost pozorování, označena  $C$  a její diskrétní úrovně jsou označeny  $C_1, C_2, \dots, C_q$ , kde  $q$  je počet skupin. Vektor nezávislých proměnných (kriterií) je označen  $\mathbf{g}$ . Vzorek pozorování, které se použijí pro tvorbu klasifikačního modelu, se

---

<sup>3</sup> Existují i metody ověřující, zda klasifikovaný objekt nepatří do nové, dosud neznámé skupiny (Ajvazjan et al. 1981; Rao 1978)

označuje jako soubor vzorů (též trénovací soubor či referenční soubor). Jednotlivá pozorování (vzory) jsou vyjádřeny pomocí vektorů  $\mathbf{x}$ , obsahující naměřené (či jinak získané) hodnoty kritérií. V trénovací fázi tvorby je pak odvozen klasifikační model  $f(\mathbf{g}) \rightarrow \hat{C}$ . V testovací fázi je ověřena míra shody apriorních a odhadnutých klasifikací,  $\hat{C} \cong C$ ; pokud je dostatečná, může být model použit pro klasifikaci nových pozorování (Doumpos a Zopounidis 2002).

Při tvorbě a hodnocení klasifikačního modelu je důležité zaměřit se na jeho přesnost, rychlost, robustnost, složitost a interpretovatelnost. Opodstatněná je snaha o co největší zjednodušení definitivního matematického modelu řešené úlohy. Čím menší počet znaků postačí pro klasifikaci, tím snazší je objasnění mechanismu sledovaného jevu, a tím názornější a srozumitelnější je interpretace získaných výsledků. Pokud se podaří snížit dimenzi na  $p = 2$ , případně  $p = 3$ , lze provést bezprostřední geometrickou interpretaci a vizuální analýzu (Ajvazjan et al. 1981).

V této kapitole budou dále popsány vybrané klasifikační metody, a to:

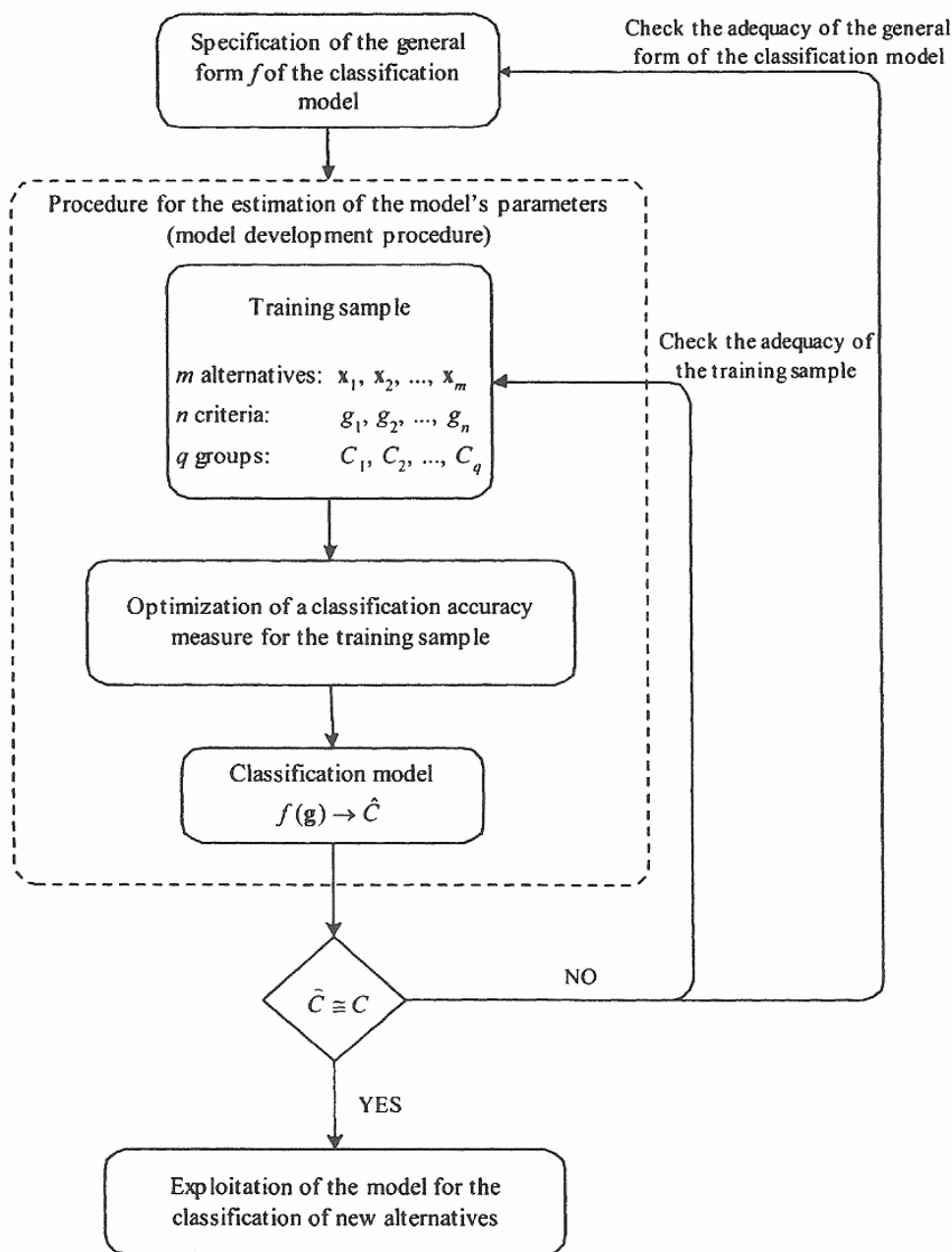
- profilová analýza;
- vícerozměrná diskriminační analýza (kanonická, lineární, kvadratická, včetně krokového výběru proměnných a robustních odhadů parametrů, neparametrické odhady hustot pomocí jádrových odhadů, nejbližších sousedů a prototypů);
- logistická a probitová regrese;
- neuronové sítě (vícevrstvé perceptronové sítě);
- klasifikační stromy a lesy.

U těchto metod bude uveden základní popis, jejich předpoklady a hlavní přednosti a omezení. Všechny uvedené metody (případně jejich určité varianty) jsou teoreticky vhodné pro predikci finanční tísně.

Pro řešení problému predikce finanční tísně lze kromě uvedených metod využít například teorii hrubých množin, support vector machines, multiple adaptive regression splines, další typy neuronových sítí (např. pravděpodobnostní) a samozřejmě k uvedeným metodám existuje řada jejich variant a úprav.

Zkoumáním, zda (či do jaké míry) jsou klasifikační metody vhodné i v praxi (tedy sestavení klasifikačního modelu s určitou vysvětlovanou proměnnou, s určitými vysvětlujícími proměnnými a pro určitý vzorek) se zabývá řada autorů a vědeckých prací, v oblasti podnikových financí např. Altman et al. (1994) (srovnání diskriminační analýzy a vícevrstvých perceptronových sítí) či Yang et al. (1997) (srovnání diskriminační analýzy, logistické regrese a neuronových sítí).

Obrázek 2 Schéma tvorby klasifikačního modelu



Pramen: Doumpos a Zopounidis (2002)

### 3.3. Profilová analýza

U jednorozměrného klasifikačního modelu je optimální hranice stanovena pro každou proměnnou a pro každou proměnnou je provedena klasifikační procedura - porovnání naměřené hodnoty se stanovenou hranicí. Tato metoda je velice jednoduchá a její aplikace nevyžaduje znalosti statistických metod. V jedné z prvních prací věnovaných předpovědi úpadku takto postupuje Beaver (1966), který metodu nazývá profilovou analýzou. Při aplikaci několika jednorozměr-

ných klasifikátorů je nevýhodou možné střídavé zařazování daného objektu do několika skupin. Tuto nevýhodu lze potlačit zavedením určitého hlasovacího pravidla mezi klasifikátory.

### 3.4. Vícerozměrná diskriminační analýza

Problémem souvislosti skupiny kvantitativních a jedné alternativní či vícehodnotové nominální proměnné se zabývá diskriminační analýza. Primárně bylo úlohou diskriminační analýzy zkoumat schopnost sledovaných proměnných přispět k odlišení jednotlivých skupin jednotek v souboru (jak ji formulovat v roce 1936 Fisher v úloze pro rozlišení třech skupin kosatců). Uvedenou souvislost lze chápat i jako pravidlo, vedoucí k zařazení jednotek do skupin na základě zjištěných hodnot několika kvantitativních proměnných. Diskriminační analýza pak směřuje ke klasifikaci jednotek s neznámou skupinovou příslušností.

#### 3.4.1. Kanonická diskriminační analýza

Základem Fisherova pojetí diskriminační analýzy je nalézt takovou lineární kombinaci  $p$  sledovaných proměnných, tedy  $\mathbf{Y} = \mathbf{b}^T \mathbf{x}$ , kde  $\mathbf{b}^T = [b_1, b_2, \dots, b_p]$  je vektor parametrů, aby lépe než jakákoliv jiná lineární kombinace separovala uvažovaných  $G$  skupin v tom smyslu, že její vnitroskupinová variabilita bude co nejmenší a meziskupinová variabilita co největší (Hebák 2004). Celková variabilita sledovaných proměnných  $\mathbf{T}$  je rovna

$$\mathbf{T} = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{ig} - \bar{x})(x_{ig} - \bar{x})^T.$$

Celkovou variabilitu lze rozložit na součet matice  $\mathbf{W}$  vyjadřující vnitroskupinovou variabilitu

$$\mathbf{W} = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)(x_{ig} - \bar{x}_g)^T,$$

a matice  $\mathbf{B}$  vyjadřující meziskupinovou variabilitu

$$\mathbf{B} = \sum_{g=1}^G \sum_{i=1}^{n_g} (\bar{x}_g - \bar{x})(\bar{x}_g - \bar{x})^T = \sum_{g=1}^G n_g (\bar{x}_g - \bar{x})(\bar{x}_g - \bar{x})^T,$$

tedy  $\mathbf{W} + \mathbf{B} = \mathbf{T}$ . Součty čtverců  $Q_B(Y)$  a  $Q_W(Y)$  představují míru meziskupinové a vnitroskupinové variability pro novou veličinu  $Y$  a lze je zapsat jako  $Q_B(Y) = \mathbf{b}^T \mathbf{B} \mathbf{b}$  a  $Q_W(Y) = \mathbf{b}^T \mathbf{W} \mathbf{b}$ . Nejvyšší meziskupinové a nejnižší vnitroskupinové variability veličiny  $Y$  se dosáhne při maximálním podílu

$$F = \frac{Q_B(Y)}{Q_W(Y)} = \frac{\mathbf{b}^T \mathbf{B} \mathbf{b}}{\mathbf{b}^T \mathbf{W} \mathbf{b}}.$$

Tento podíl se nazývá Fisherovo diskriminační kritérium (Hebák 2004). Pro stanovení veličiny  $Y = \mathbf{b}^T \mathbf{x}$ , která by postihovala odlišnosti mezi skupinami, je třeba určit prvky vektoru  $\mathbf{b}$  tak, aby maximalizoval diskriminační kritérium. Maximalizační úlohu lze řešit

$$(\mathbf{B} - \lambda \mathbf{W}) \mathbf{b} = 0,$$

resp.

$$(\mathbf{W}^{-1} \mathbf{B} - \lambda \mathbf{I}) \mathbf{b} = 0 \quad |\mathbf{W}| \neq 0.$$

Charakteristická rovnice  $|\mathbf{W}^{-1} \mathbf{B} - \lambda \mathbf{I}| = 0$  má  $r$  řešení, kterými jsou charakteristická čísla  $\lambda_1, \lambda_2, \dots, \lambda_r$  matice  $\mathbf{W}^{-1} \mathbf{B}$ . Největšímu z těchto charakteristických čísel  $\lambda_1$  odpovídá charakteristický vektor  $\mathbf{b}_1$ , který maximalizuje diskriminační kritérium  $F$ . Charakteristická rovnice neurčuje vektor  $\mathbf{b}_1$  jednoznačně, ale pouze stanovuje poměr mezi jeho prvky. Konkrétní hodnoty prvků vektoru  $\mathbf{b}_1$ , tedy koeficienty pro hledanou lineární kombinaci je vhodné volit tak, aby

$$\frac{1}{n-G} \mathbf{b}_1^T \mathbf{W} \mathbf{b}_1 = 1,$$

tedy vydělením vektoru  $\mathbf{b}_1$  výrazem

$$\sqrt{\frac{\mathbf{b}_1^T \mathbf{W} \mathbf{b}_1}{n-G}}.$$

Potom vnitroskupinovou variabilitu veličiny  $Y_1 = \mathbf{b}_1^T \mathbf{x}$  vyjadřuje jednotkový rozptyl a kritérium  $F$  lze zapsat jako

$$F = \frac{1}{n-G} \mathbf{b}_1^T \mathbf{B} \mathbf{b}_1$$

a příslušné charakteristické číslo vyjadřuje míru meziskupinové variability veličiny  $Y_1$ .

Je-li soubor jednotek popsaných  $p$  proměnnými tříděn do dvou skupin, stačí pro vyjádření celkové variability původních proměnných jediný diskriminant (Hebák 2004). V případě třídění do více než dvou skupin lze však prostřednictvím jednoho diskriminantu vyjádřit pouze část variability původních proměnných. Použitím dalších charakteristických čísel  $\lambda_2, \lambda_3, \dots, \lambda_r$  a charakteristických vektorů  $\mathbf{b}_2, \mathbf{b}_3, \dots, \mathbf{b}_r$  obdržíme další kanonické proměnné  $Y_j = \mathbf{b}_j^T \mathbf{x}$ ,  $j = 2, 3, \dots, r$ . Tyto diskriminanty jsou vzájemně nezávislé, jejich maximální počet je dán výrazem  $r = \min(p, G - 1)$ .

Prvky vektoru  $\mathbf{b}_j$ ,  $\mathbf{b}_j^T = [\mathbf{b}_{j1}, \mathbf{b}_{j2}, \dots, \mathbf{b}_{jp}]$  jsou koeficienty  $j$ -té kanonické proměnné. Dosadíme-li do  $Y_j = \mathbf{b}_j^T \mathbf{x}$  pro každou jednotku zjištěné hodnoty veličin  $X_1, X_2, \dots, X_p$ , získáme její diskriminační skóre. Použijeme-li při stanovení skóre konstantu  $c_j$

$$c_j = -\mathbf{b}_j^T \bar{\mathbf{x}},$$

pak je průměrné diskriminační skóre jednotlivých diskriminantů nulové. Pro  $i$ -tou jednotku ( $i = 1, 2, \dots, n_g$ ) v  $g$ -té skupině ( $g = 1, 2, \dots, G$ ) určíme  $j$ -té diskriminační skóre ( $j = 1, 2, \dots, r$ ) jako

$$y_{ijg} = c_j + \sum_{k=1}^p b_{jk} x_{igk}.$$

Představu o tom, jak se z hlediska  $j$ -té kanonické proměnné skupiny od sebe liší, lze získat výpočtem průměrných hodnot diskriminantů ve skupinách (skupinových centroidů)

$$\bar{y}_{jg} = c_j + \sum_{k=1}^p b_{jk} \bar{x}_{gk}.$$

Pro interpretaci výsledků diskriminační analýzy se často koeficienty normují vynásobením směrodatnými odchylkami vyjadřujícími míru vnitroskupinové variability původní veličiny (Hebák 2004). Označíme-li  $\mathbf{F}$  diagonální matici s odmocninami diagonálních prvků matice  $\mathbf{W}$ , pak normované koeficienty dostáváme jako

$$\mathbf{b}^* = \frac{1}{\sqrt{n-G}} \mathbf{F} \mathbf{b}.$$

Alternativní přístup k interpretaci používá korelační koeficienty mezi kanonickou proměnnou a původními proměnnými (tzv. strukturní koeficienty). Jejich vysoká kladná či záporná hodnota vypovídá o tom, že sledovaná proměnná je pro daný diskriminant charakteristická. Ze znaménka vyplývá, zda vyšší hodnoty původní proměnné vedou ke zvýšení či snížení hodnoty diskriminačního skóre. Vektor těchto korelačních koeficientů pro  $j$ -tý diskriminant:

$$a_j = \frac{1}{\sqrt{n-G}} \mathbf{F}^{-1} \mathbf{W} \mathbf{b}.$$

Diskriminační skóre lze využít pro klasifikaci  $n$  objektů do  $G$  skupin. Objekt bude zařazen do té skupiny, která je nejbližší ve smyslu vzdálenosti objektu od skupinového centroidu. Diskriminanty jsou nekorelované, vzdálenost objektu od centroidu lze (pro  $r$  diskriminantů) vyjádřit euklidovskou vzdáleností:

$$d_{ig}^2 = \sum_{j=1}^r (y_{ij} - \bar{y}_{jg})^2 = \sum_{j=1}^r [b_j^T (x_i - \bar{x}_g)]^2.$$

Objekt pak bude zařazen do skupiny, pro níž je  $d_{ig}^2$  nejmenší.

### 3.4.1.1. Test významnosti diskriminantů

Test slouží k určení diskriminantů, které je vhodné použít pro odlišení jednotlivých skupin a které lze pominout. Diskriminační kritérium  $\lambda = \mathbf{b}^T \mathbf{B} \mathbf{b} / \mathbf{b}^T \mathbf{W} \mathbf{b}$  nabývá nejvyšší hodnoty  $\lambda_1$  (nejvyšší vlastní číslo matice  $\mathbf{W}^{-1} \mathbf{B}$ ). Test o shodě vektorů středních hodnot v  $G$  skupinách je založen na Wilksově statistice  $\Lambda$ ,

$$\Lambda_1 = \prod_{i=1}^r \frac{1}{1 + \lambda_i},$$

kteřá má rozdělení  $\Lambda_{p, G-1, n-G}$ . Testovací kritérium

$$V_1 = \left[ n - 1 - \frac{1}{2}(p + G) \right] \sum_{i=1}^r \ln(1 + \lambda_i)$$

má přibližně  $\chi^2$  rozdělení s  $p(G-1)$  stupni volnosti. Test hypotézy o shodě vektorů skupinových středních hodnot je ekvivalentní testu hypotézy  $\lambda_1 = \lambda_2 = \dots = \lambda_r = 0$ . Zamítnutí této hypotézy znamená, že nejméně jedno z vlastních čísel (tedy  $\lambda_1$ , které je největší) je nenulové a diskriminační kritérium významné. Navazuje ověření hypotézy  $\lambda_2 = \lambda_3 = \dots = \lambda_r = 0$ , kdy testovací kritérium je

$$V_2 = \left[ n - 1 - \frac{1}{2}(p + G) \right] \sum_{i=2}^r \ln(1 + \lambda_i),$$

kteřé má  $\chi^2$  rozdělení s  $(p-1)(G-2)$  stupni volnosti. Pokud test vede k zamítnutí nulové hypotézy, předpokládáme, že  $\lambda_2$  je významné. Dále pokračujeme testováním každého  $\lambda_i$  až k testu, kdy se nulovou hypotézu nepodaří zamítnout. Obecně má testovací kritérium v  $m$ -tém kroku

$$\Lambda_m = \prod_{i=m}^r \frac{1}{1 + \lambda_i}$$

rozdělení  $\Lambda_{p-m+1, G-m, n-G-m+1}$ . Statistika

$$V_m = \left[ n - 1 - \frac{1}{2}(p + G) \right] \sum_{i=m}^r \ln(1 + \lambda_i)$$

má  $\chi^2$  rozdělení s  $(p-m+1)(G-m)$  stupni volnosti (Hebák 2004; Rencher 2002).

### 3.4.1.2. Krokový výběr proměnných

Často je k dispozici velký počet znaků a je nutné rozhodnout, zda a které je možné pro separaci do skupin vynechat. Při zaváděcí krokové metodě (forward selection) se začíná s jedním znakem, který separuje skupiny maximálně. V každém dalším kroku se přidává ten znak, který nejvíce přispívá k separaci. Zpětná kroková metoda (backward elimination) je podobný postup, při kterém se začíná se všemi znaky a v každém kroku je odebrán znak, který přispívá k separaci skupin nejméně. Krokový výběr (stepwise selection) je kombinací obou postupů. V každém kroku je po přidání proměnné vyhodnoceno, zda některá již dříve přidaná proměnná není redundantní vzhledem k nově přidané proměnné. Nejlepší podskupina je takový postup, kdy jsou prověřeny všechny kombinace znaků a vybrána je ta kombinace, která skupiny separuje nejlépe (Rencher 2002).



### 3.4.1.3. Krokový výběr proměnných na základě Wilksova kritéria

Krokový výběr proměnných je založen na vícerozměrné jednofaktorové analýze rozptylu. Předpokladem je, že výběr pochází z vícerozměrného normálního rozdělení se shodnými kovariančními maticemi v jednotlivých skupinách (Rencher 2002). Hodnota Wilksova kritéria je pak dána

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|}.$$

Nulovou hypotézu  $H_0: \mu_1 = \mu_2 = \dots = \mu_G$  zamítáme, pokud  $\Lambda \leq \Lambda_{\alpha, p, n-G, G-1}$ . Statistiku  $\Lambda$  je možné aproximovat statistikou  $F$ , která je dána

$$F = \frac{1 - \Lambda}{\Lambda} \frac{n - G - p}{G - 1}$$

a má přibližně  $F$  rozdělení s  $G - 1$  a  $n - G$  stupni volnosti. Stupně volnosti zůstávají ve všech krocích stejné. V druhém a dalších krocích se užívá parciální  $\Lambda$ , která je dána

$$\Lambda = \frac{\Lambda_{p+1}}{\Lambda_p},$$

kde  $\Lambda_p$  je hodnota Wilksova kritéria před přidáním proměnné a  $\Lambda_{p+1}$  je Wilksova lambda po přidání proměnné do modelu. V každém kroku je pak

1. Proměnná s nejnižší hodnotou  $F$  odstraněna z modelu, pokud její hodnota není větší či rovna  $F_{\text{out}}$ .
2. Proměnná s nejvyšší hodnotou  $F$  zařazena do modelu, pokud její hodnota není menší než  $F_{\text{in}}$ .
3. Procedura končí, pokud v daném kroku nelze žádnou proměnnou zařadit či vyřadit.

### 3.4.2. Klasifikační diskriminační analýza

Předpokládejme, že populace je rozdělena do dvou skupin a že rozdělení vícerozměrné náhodné veličiny  $\mathbf{x}$  ve dvou skupinách je vícerozměrné normální s vektory středních hodnot  $\boldsymbol{\mu}_1$  a  $\boldsymbol{\mu}_2$  a shodnými kovariančními maticemi  $\boldsymbol{\Sigma}$  (Hebák 2004). Charakteristický vektor  $\mathbf{b}$ , který maximalizuje Fisherovo diskriminační kritérium  $F$  lze vyjádřit jako

$$\mathbf{b} = k\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad \text{a} \quad \frac{1}{k}\mathbf{b} = \mathbf{a} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2),$$

kde  $k$  je libovolná konstanta. Pokud  $\mathbf{b}^T \mathbf{W} \mathbf{b} / (n - G) = 1$ , je

$$k = [(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^{1/2} \quad \text{a} \quad k > 0.$$

Diskriminační funkce

$$\mathbf{x}^T \mathbf{a} = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

má ve skupinách normální rozdělení se středními hodnotami

$$\mathbf{x}_1^T \mathbf{a} = \mathbf{x}_1^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad \text{a} \quad \mathbf{x}_2^T \mathbf{a} = \mathbf{x}_2^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Jejich vzdálenost

$$(\boldsymbol{\mu}_1^T \mathbf{a} - \boldsymbol{\mu}_2^T \mathbf{a}) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

je Mahalanobisovou vzdáleností dvou skupin. Střed mezi těmito středními hodnotami je roven

$$c = \frac{1}{2}(\boldsymbol{\mu}_1^T \mathbf{a} + \boldsymbol{\mu}_2^T \mathbf{a}) = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Pokud je  $\mathbf{x}^T \mathbf{a} > c$ , má klasifikovaná jednotka blíže k první skupině, v opačném případě má blíže k druhé skupině.

Uvedená kritéria však apriorně předpokládají stejné zastoupení obou skupin v populaci, a tedy stejnou pravděpodobnost mylného zařazení objektu z první skupiny do druhé a naopak. Rozsahy skupin se však mohou lišit. Tvoří-li jednu skupinu  $100\pi_1\%$  a druhou  $100(1 - \pi_1)\%$  objektů populace, je to významná informace pro klasifikaci jednotek neznámého původu už v okamžiku, kdy o nich žádné jiné informace nemáme (Hebák 2004).

Předpokládejme, že  $\pi_1$  a  $\pi_2$  jsou rozsahu skupiny odpovídající apriorní pravděpodobnosti příslušnosti objektu k určité skupině. Na základě hodnot  $p$  znaků zjištěných u určitého objektu lze uvažovat podmíněnou, aposteriorní pravděpodobnost této příslušnosti, kterou lze podle Bayesova vzorce vyjádřit

$$\frac{\pi_g f_g(\mathbf{x})}{\pi_1 f_1(\mathbf{x}) + \pi_2 f_2(\mathbf{x})} \quad \text{pro } g = 1, 2.$$

Objekt neznámého původu bude klasifikován do skupiny s nejvyšší aposteriorní pravděpodobností, tedy např. do 1. skupiny, pokud  $\pi_1 f_1(\mathbf{x}) > \pi_2 f_2(\mathbf{x})$ . Klasifikační pravidlo pro zařazení do 1. skupiny je tedy

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} > \frac{\pi_2}{\pi_1},$$

v opačném případě do 2. skupiny.

Pokud obě chyby klasifikace mají stejnou váhu, je optimální rozhodovací pravidlo, které minimalizuje celkovou pravděpodobnost mylné klasifikace. Pokud mají chyby různou váhu, užije se ztrátová matice:

$$Z = \begin{bmatrix} 0 & z(2|1) \\ z(1|2) & 0 \end{bmatrix}$$

Ztráta může být

- nulová, při správné klasifikaci;
- $z(1|2)$ , pokud je pozorování ze skupiny 2 mylně zařazeno do skupiny 1;
- $z(2|1)$ , pokud je pozorování ze skupiny 1 chybně zařazeno do skupiny 2.

a optimální je pak postup, který minimalizuje celkovou ztrátu. Objekt bude zařazen do první skupiny, pokud

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} > \frac{z(1|2)\pi_2}{z(2|1)\pi_1}.$$

Předpokládejme, že skupina  $g_1$  s hustotou  $f_1(\mathbf{x})$  má normální rozdělení  $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  a skupina  $g_2$  s hustotou  $f_2(\mathbf{x})$  má normální rozdělení  $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ . Diskriminační analýzu můžeme dělit na lineární a kvadratickou. O lineární diskriminační analýze mluvíme v případě, že se hustoty  $f_1(\mathbf{x})$  a  $f_2(\mathbf{x})$  liší pouze středními hodnotami (a tedy  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ ). Pokud je odlišnost daná navíc kovariančními maticemi, hovoříme o kvadratické diskriminační analýze.

### 3.4.2.1. Lineární diskriminační analýza

Předpokládejme, že hustota  $f_i(\mathbf{x})$  náhodného výběru  $X$  pro skupinu  $g_i$ , kde  $i = 1, 2$ , je dána vztahem

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right].$$

Objekt  $\mathbf{x}$  je pak zařazen do skupiny  $g_1$ , pokud

$$\mathbf{x}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) > \ln \frac{z(1|2)\pi_1}{z(2|1)\pi_2},$$

v opačném případě je objekt  $\mathbf{x}$  zařazen do skupiny  $g_2$ .

### Technika diskriminace do více skupin

Za předpokladu vícerozměrné normality a shody kovariančních matic lze lineární diskriminační kritérium použité pro dvě skupiny rozšířit pro případ více skupin. Toto kritérium se počítá pro každou třídu zvlášť (Meloun a Militký 2006). Při klasifikaci objektů se pak objekt zařazuje do té třídy, pro kterou vyjde hodnota lineárního diskriminačního kritéria nejvyšší. Lineární diskriminační kritérium pro  $g$ -tou skupinu má tvar

$$LDK_g(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_g - \frac{1}{2} \boldsymbol{\mu}_g^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_g + \ln \pi_g$$

Neznámé vektory středních hodnot je přitom třeba odhadnout vektory výběrových průměrů  $\bar{\mathbf{x}}_g$  a kovarianční matici společnou výběrovou kovarianční maticí  $\mathbf{S}$ , apriorní pravděpodobnosti  $\pi_g$  jsou odhadnuty výběrovými podíly jednotlivých skupin  $p_g$ .

### 3.4.2.2. Kvadratická diskriminační analýza

Pokud se hustoty  $f_1(\mathbf{x})$  a  $f_2(\mathbf{x})$  liší jak středními hodnotami, tak i kovariančními maticemi, pak se jedná o kvadratickou diskriminační analýzu. Objekt  $\mathbf{x}$  je pak zařazen do skupiny  $g_1$ , pokud

$$\frac{1}{2} \mathbf{x}^T (\boldsymbol{\Sigma}_2^{-1} - \boldsymbol{\Sigma}_1^{-1}) \mathbf{x} + (\boldsymbol{\mu}_1^T \boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}_2^{-1}) \mathbf{x} + \frac{1}{2} \ln \frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|} - \frac{1}{2} (\boldsymbol{\mu}_1^T \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\mu}_2) > \ln \frac{z(1|2)\pi_1}{z(2|1)\pi_2}$$

v opačném případě je objekt  $\mathbf{x}$  zařazen do skupiny  $g_2$ .

### Technika diskriminace do více skupin

Lze definovat kvadratické diskriminační kritérium

$$QDK_g(\mathbf{x}) = -\frac{1}{2} \ln |\boldsymbol{\Sigma}_g| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\mathbf{x} - \boldsymbol{\mu}_g) + \ln \pi_g$$

Objekt  $\mathbf{x}$  se pak zařazuje do té skupiny, které odpovídá maximální hodnota  $QDK_g(\mathbf{x})$ .

### 3.4.3. Porušení předpokladů

Diskriminační analýza je založena na splnění několika předpokladů:

1. soubor je rozdělen do skupin, přičemž každé pozorování patří právě do jedné skupiny;
2. rozdělení sledovaných proměnných je vícerozměrné normální;
3. jsou známy apriorní pravděpodobnosti a náklady mylné klasifikace;
4. pro lineární diskriminační analýzu jsou kovarianční matice ve skupinách shodné;
5. jednotlivé proměnné nejsou vzájemně korelované.

Hypotézu o shodě empirického rozdělení s vícerozměrným normální rozdělením ověřujeme pomocí testů založených na vícerozměrné šikmosti a vícerozměrné špičatosti Hebač et al. (2004). Koeficient vícerozměrné šikmosti  $B_1$  má tvar

$$B_1 = 1/n^2 \sum_{i=1}^n \sum_{i=1}^n [(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})]^3.$$

Odchylka od hodnoty šikmosti vícerozměrného normálního rozdělení je pokládána na hladině  $\alpha$  za významnou, pokud

$$\frac{n}{6} B_1 \geq \chi_{1-\alpha/2}^2 \left( \frac{p(p+1)(p+2)}{6} \right).$$

Koeficient vícerozměrné špičatosti  $B_2$  má tvar

$$B_2 = 1/n \sum_{i=1}^n [(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})]^2.$$

Odchylka od hodnoty špičatosti vícerozměrného normálního rozdělení je pokládána na hladině  $\alpha$  za významnou, pokud

$$\sqrt{\frac{n}{8p(p+2)}} [B_2 - p(p+2)] \geq u_{1-\alpha/2}.$$

Podmínka vícerozměrné normality je u finančních ukazatelů často porušena. Obvykle se lze setkat s postupem, kdy jsou proměnné testovány na jednorozměrnou normalitu a při jejím porušení jsou užity některé transformační metody či metody odstranění odlehlých pozorování. Micha (1984) odstraňuje v každém roce 10 až 12 % pozorování, a to podle celkové euklidovské vzdálenosti standardizovaných ukazatelů  $d_j^2$ ,

$$d_j^2 = \sum_i \left( \frac{x_{ij} - \bar{x}_i}{\sigma_i} \right)^2,$$

kde  $x_{ij}$  je hodnota ukazatele  $i$  pro podnik  $j$ ,  $\bar{x}_i$  je průměrná hodnota ukazatele  $i$  a  $\sigma_i$  je směrodatná odchylka ukazatele  $i$ . Taffler (1982) nahrazuje pozorování vzdálená více než 4 směrodatné odchylky od průměru zbylých pozorování tímto průměrem; pozorování vzdálená  $2,5 \sigma$  až  $4 \sigma$  nahrazuje hodnotou odpovídající hranici  $2,5 \sigma$ .

Zde může dojít k těmto chybám:

1. Jednorozměrná normalita není postačující podmínka pro vícerozměrnou normalitu. Drobné odchylky od normality vícerozměrné analýze obvykle nebrání, s rostoucím rozsahem výběru působí jako překážka méně. Testy normality lze použít orientačně, než se doslova řídit jejich závěrem (Hebák et al. 2004).
2. Transformace může změnit vzájemné vztahy mezi proměnnými. Zde se jedná o nelineární transformace, neboť lineární transformace by zachovala původní tvar rozdělení. Aplikaci statistických metod na transformovaná data je třeba vzít v úvahu při interpretaci výsledků. Pro přiblížení dat normalitě se užívá transformace: a) odmocninová; b) logitová; c) logaritmická (Hebák et al. 2004).
3. Odstranění odlehlých pozorování může vést ke ztrátě informace. Střední vektory jsou odhadnuty pomocí robustních metod, které snižují či zcela eliminují vliv odlehlých pozorování, např. MCD-odhady (Minimum Covariance Determinant) a MVE-odhady (Minimum Volume of Ellipsoid) (Hubert a Van Driessen 2004; Rousseeuw a Van Driessen 1999).

V případě, že kovarianční matice nejsou shodné, by měla být aplikována kvadratická diskriminační analýza. Předpoklad shodných kovariančních matic však finanční ukazatele splňují jen vzácně. Lineární diskriminační analýza je v některých případech proti tomuto porušení robustní, kvadratickou diskrimi-

nační analýzu je však nutnou použít v případě souborů velkého rozsahu, malého počtu nezávislých proměnných a velkých rozdílů mezi kovariančními maticemi.

Vynechání apriorních pravděpodobností a nákladů mylné klasifikace vede k chybnému odhadu pravděpodobnosti správné klasifikace. Ve většině studií jsou zanedbány i tyto parametry, výjimkou je např. Altman et al. (1977) – podrobněji v kap. 4 Klasifikační modely. Důvodem může být i to, že jejich odhady jsou obtížné a do značné míry subjektivní, proto do modelů vstupují shodné náklady mylné klasifikace a shodné apriorní pravděpodobnosti. V modelech predikce finanční tísně lze vyjít z příkladu banky, která hodnotí žadatele před poskytnutím úvěru a kde se náklady chyby I. druhu liší od nákladů chyby II. druhu. V prvním případě mohou náklady mylné klasifikace činit až 100 % výše úvěru, v druhém případě se jedná o oportunitní náklady – ušlý zisk (Agarwal a Taffler 2007). Gentry et al. (1985) uvádějí, že v letech 1970 až 1981 byla v USA průměrná míra úpadku 0,038 %.

Požadavek absence multikolinearity je často považován za bezvýznamný. Obecně však platí, že silná korelace mezi proměnnými vede k nestabilním a obtížně interpretovatelným koeficientům diskriminační funkce. Faktem je, že poměrové ukazatele užívané v modelech predikce finanční tísně jsou často založené na shodných jmenovatelích či čitatelích.

Kvadratické diskriminační kritérium je značně citlivé na odchylky od normality, proto se v případech, kdy jsou rozdíly mezi kovariančními maticemi malé, doporučuje používat spíše lineární diskriminační kritérium. Lineární diskriminační kritérium je proti odchýlkám od normality poměrně robustní, pokud jsou však vysvětlující proměnné např. binární, je vhodné použít jiné metody – neparametrické. Při porušení předpokladu rovnosti kovariančních matic poskytuje kvadratické diskriminační kritérium výrazně lepší výsledky než lineární diskriminační kritérium v případě výběrů velkého rozsahu a velkých rozdílů mezi kovariančními maticemi. Pro výběry malého rozsahu a s malým rozdílem mezi kovariančními maticemi je však kvadratické diskriminační kritérium výrazně horší než lineární (Havránek a Vorlíček 1980).

#### 3.4.4. Robustní MCD-odhady

MCD-odhad pro  $j$ -tou skupinu je definovaný jako střední hodnota  $\hat{\boldsymbol{\mu}}_{j,0}$  a kovarianční matice  $\mathbf{S}_{j,0}$  pro  $h_j$  pozorování ( $z n_j$ ), jejichž kovarianční matice má nejmenší determinant. Množství pozorování  $h_j$  by mělo být větší než  $[(n_j + p + 1) / 2]$  a  $n_j - h_j < v$ , kde  $v$  je počet odlehlých pozorování v  $j$ -té skupině. Protože  $v$  je většinou neznámé, bereme za  $h_j$  pevně zvolenou konstantu  $h_j = [(n_j + p + 1) / 2]$ . Pokud však předpokládáme například méně než 25 % odlehlých pozorování uvnitř každé skupiny, doporučuje se vzít  $h_j \approx 0,75 n_j$  (Horáková 2008, Hubert a Van Driessen 2002). Na základě výchozích odhadů  $\hat{\boldsymbol{\mu}}_{j,0}$  a  $\mathbf{S}_{j,0}$  se pro každé pozorování  $\mathbf{x}_{ij}$  ze skupiny  $j$  určí jejich předběžné robustní vzdálenosti,

$$RD_{ij}^0 = \sqrt{(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_{j,0})^T \mathbf{S}_{j,0}^{-1} (\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_{j,0})}.$$

Pro každé pozorování ze skupiny  $j$  se položí

$$w_{ij} = 1, \text{ pokud } RD_{ij}^0 \leq \sqrt{\chi_{0,975}^2(p)}, \text{ jinak } w_{ij} = 0.$$

O MCD-odhadu se hovoří tehdy, pokud se docílí odhadů střední hodnoty  $\hat{\boldsymbol{\mu}}_{j,MCD}$  a kovarianční matice  $\mathbf{S}_{j,MCD}$  ve skupině  $j$  z pozorování s váhou 1. Robustní odhady polohy a rozptylu umožňují označit odlehlá pozorování. Nejprve se určí tzv. konečná Mahalanobisova vzdálenost pro každé pozorování  $\mathbf{x}_{ij}$  ze skupiny  $j$  od  $\hat{\boldsymbol{\mu}}_{j,MCD}$

$$RD_{ij} = \sqrt{(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_{j,MCD})^T \boldsymbol{\Sigma}_{j,MCD}^{-1} (\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_{j,MCD})}.$$

Pak se  $\mathbf{x}_{ij}$  považuje za odlehlé pozorování, pokud  $RD_{ij} > \sqrt{\chi_{0,975}^2(p)}$ . Označme  $\tilde{n}_j$  počet „neodlehlých“ pozorování ve skupině  $j$  a  $\tilde{n} = \sum_j \tilde{n}_j$ , pak robustní odhad apriorní pravděpodobnosti příslušnosti daného objektu do skupiny  $j$  je dán vztahem

$$\pi_j^R = \frac{\tilde{n}_j}{\tilde{n}}.$$

Robustní kvadratické diskriminační kritérium je tedy: objekt  $\mathbf{x}$  je zařazen do skupiny, pro níž je hodnota  $RQDK_j(\mathbf{x})$  nejvyšší,

$$RQDK_j(\mathbf{x}) = -\frac{1}{2} \ln |\hat{\boldsymbol{\Sigma}}_{j,MCD}| - \frac{1}{2} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,MCD})^T \hat{\boldsymbol{\Sigma}}_{j,MCD}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,MCD}) + \ln \pi_j^R.$$

Pro lineární případ je potřebný odhad společné kovarianční matice  $\boldsymbol{\Sigma}$ , jenž je možno obdržet třemi postupy. První metoda spočívá ve sdružení robustních kovariančních matic jednotlivých tříd, zatímco druhá metoda sdružuje pozorování. Třetí přístup se nazývá „minimum within-group covariance determinant“ (Horáková 2008).

Robustní lineární diskriminační kritérium je pak: objekt  $\mathbf{x}$  je zařazen do skupiny, pro níž je hodnota  $RLDK_j(\mathbf{x})$  nejvyšší,

$$RLDK_j(\mathbf{x}) = \mathbf{x}^T \hat{\boldsymbol{\Sigma}}_{MCD}^{-1} \hat{\boldsymbol{\mu}}_{j,MCD} - \frac{1}{2} \hat{\boldsymbol{\mu}}_{j,MCD}^T \hat{\boldsymbol{\Sigma}}_{MCD}^{-1} \hat{\boldsymbol{\mu}}_{j,MCD} + \ln \pi_j^R.$$

### 3.4.5. Jádrové odhady hustoty

Lineární a kvadratická diskriminační analýza popsaná v předchozí části byla založena na předpokladu, že data pocházejí z vícerozměrného normálního rozdělení. Pozorování  $y$  je pak zařazeno do skupiny, pro kterou je  $p_i f(y|G_i)$  maximální. Pokud je rozdělení veličiny nenormální či neznámé, hustota pravděpodobnosti může být odhadnuta na základě jádrových odhadů hustoty (Rencher 2002). Jádrem se rozumí libovolná funkce, která (Meloun a Militký 2006):

- je nezáporná  $K(y) \geq 0$ ;
- je symetrická kolem nuly  $K(y) = K(-y)$ ;
- má vlastnosti hustoty pravděpodobnosti.

Přehled nejdůležitějších jader uvádí např. Orava (2006), a to Gaussovo, Laplaceovo, Cauchyovo, Epanechnikovo, trojúhelníkové, obdélníkové, kosinové či kvartické. Jádrový odhad hustoty je definován vztahem

$$\hat{f}(y) = \frac{1}{hn} \sum_{i=1}^n K\left(\frac{y - y_i}{h}\right).$$

Parametr  $h$  se nazývá šířka okna (vyhlazovací parametr, bandwidth) a významně ovlivňuje kvalitu odhadu. Pro odhad parametru šířky okna je odvozeno několik metod (Orava 2008). Pro odhad hustoty pravděpodobnosti náhodných vektorů jsou definovány vícerozměrné jádrové odhady hustoty (Forbelská 2000),

$$\hat{f}(\mathbf{y}) = \frac{1}{nh_1 \dots h_p} \sum_{i=1}^n K\left(\frac{y_1 - y_{i1}}{h_1}, \dots, \frac{y_p - y_{ip}}{h_p}\right).$$

Při vlastní klasifikaci neznámých pozorování je nutné provést odhad hustoty pro každou skupinu, tedy  $\hat{f}(\mathbf{y} | G_1)$ ,  $\hat{f}(\mathbf{y} | G_2)$ , ...,  $\hat{f}(\mathbf{y} | G_k)$ , kde  $\mathbf{y}$  je vektor pozorování neznámé příslušnosti. Klasifikační pravidlo je tedy: Zařad'  $\mathbf{y}$  do skupiny  $G_i$ , pro kterou je  $\hat{f}(\mathbf{y} | G_i)$  maximální.

### 3.4.6. Metoda nejbližších sousedů

Klasifikace podle nejbližších sousedů patří mezi neparametrické metody klasifikace, u kterých nejsou vysloveny předpoklady o tvaru hustoty pravděpodobnosti. Metoda nejbližšího souseda je založena na hledání přímo aposteriorní pravděpodobnosti.

Rozlišuje se klasifikace podle nejbližšího souseda (1-NN) a klasifikace pro obecně  $k$  sousedů ( $k$ -NN). V obecném případě klasifikace podle všech prvků trénovací množiny je třeba porovnat všechny její prvky s prvkem právě klasifikovaným, časová náročnost algoritmu je tedy úměrná rozsahu trénovací množiny. Časovou náročnost lze částečně snížit vhodným setříděním prvků nebo jejich vhodným rozdělením do skupin. Při praktickém použití se chyba klasifikace  $k$ -NN blíží chybě mnohem složitějších metod, jako např. neuronových sítí; z toho důvodu se často používá jako referenční metoda (Houdek et al. 2001).

Výhodou  $k$ -NN klasifikátoru je, že není nutné znát rozdělení pravděpodobnosti zkoumaných dat. Nevýhodou je nutnost normalizace dat (při rozdílných metrickách proměnných) a paměťová náročnost.



### Postup klasifikace 1-NN

1. Výpočet vzdálenosti neznámého prvku od všech objektů trénovacího souboru.
2. Zařazení neznámého objektu do té skupiny, ve které se nachází i nejbližší objekt trénovacího souboru.

### Postup klasifikace k-NN

1. Kolem neznámého prvku se vytvoří hyperkoule, která obsahuje právě  $k$  objektů trénovacího souboru.
2. Neznámý objekt klasifikujeme do té skupiny, která je v hyperkouli zastoupena největším počtem objektů. Pokud se rozsahy skupin liší, klasifikujeme do skupiny  $i$ , pro níž je nejvyšší poměr  $k_i / n_i$  (Rencher 2002).

Při použití metod  $k$ -NN pro  $k > 1$  je nutné dobře zvolit hodnotu  $k$ . Pro jednoznačnost klasifikace se u dvou tříd volí  $k$  vždy liché. Pro více tříd mohou nastat situace, kdy nelze jednoznačně rozhodnout.

### 3.4.7. Metoda nejbližších prototypů

Při metodě nejbližších prototypů (též predikční analýza microarray (Schlesinger 2004)) jsou naměřené údaje transformovány tak, aby se zvýraznily znaky přispívající k odlišení jednotlivých tříd. Naopak znaky málo významné jsou při konstrukci klasifikátoru zanedbávány. Pro každý znak se zavádí vzdálenost průměrů v jednotlivých třídách od celkové průměrné hodnoty příslušného znaku,

$$d_{jk} = \frac{(\bar{x}_{jk} - \bar{x}_j)}{s_j}$$

kde  $\bar{x}_j$  je celkový průměr  $j$ -tého znaku,  $\bar{x}_{jk}$  je průměr  $j$ -tého znaku v rámci  $k$ -té třídy a  $s_j^2$  představuje společnou vnitroskupinovou variabilitu  $j$ -tého znaku (Tibshirani 2002; Schlesinger 2004). K porovnávání a identifikaci podstatných znaků se využívá standardizovaná mezitřídní odchylka  $d_{jk}$  a práh významnosti  $\Delta \geq 0$ . Postup výběru pro klasifikaci důležitých znaků spočívá v porovnání  $d_{jk}$  s prahovou hodnotou  $\Delta$ ,

$$d'_{jk} = \begin{cases} (\text{sgn}(d_{jk}))(|d_{jk}| - \Delta) & \text{pokud } |d_{jk}| - \Delta > 0 \\ 0 & \text{jinak} \end{cases}$$

Tento postup v závislosti na nastavení prahu  $\Delta$  zmenší vzdálenost průměrných znaků v jednotlivých třídách od příslušných celkových průměrů právě o prahovou hodnotu, resp. vzdálenost nevýznamných znaků zanedbá. Nastavení prahu  $\Delta$  ovlivňuje počet znaků, které budou považovány za podstatné, v krajních případech může postup výběru jako podstatné označit všechny zna-

ky, ale také žádný. Při  $\Delta = 0$  do klasifikační procedury vstupují všechny znaky, s jeho postupným zvyšováním se počet znaků snižuje. Pomocí křížového ověřování se určí  $\Delta$ , pro které je celková chyba modelu nejnižší (Tibshirani 2002; Schlesinger 2004). Upravené vzdálenosti  $d'_{jk}$  se využijí k transformaci průměrných hodnot jednotlivých znaků,

$$\bar{x}'_{jk} = \bar{x}_j + s_j d'_{jk}.$$

Upravené hodnoty se využijí při konstrukci klasifikátoru. Vektor transformovaných průměrných znaků  $k$ -té třídy  $\bar{\mathbf{x}}'_k = (\bar{x}'_{1k}, \dots, \bar{x}'_{pk})^T$  představuje vektor znaků průměrného objektu  $k$ -té třídy a nazývá se prototyp třídy  $k$ . Klasifikátor metody predikuje objektu  $\mathbf{x}$  třídu s nejbližším upraveným prototypem  $\bar{\mathbf{x}}'_k$ . Pro každou třídu je definován diskriminační skór,

$$\delta_k(\mathbf{x}) = \sum_{j=1}^p \frac{(x_j - \bar{x}'_{jk})^2}{s_j^2} - 2 \log \pi_k.$$

První člen je dán čtvercem standardizované vzdálenosti  $\mathbf{x}$  od  $k$ -tého prototypu a druhý člen odpovídá korekci apriorní pravděpodobností. Na základě těchto diskriminačních skóru je objekt klasifikován do skupiny, pro níž je skór minimální. Diskriminační skóry jsou podobné diskriminačním skóruům užívaným v lineární diskriminační analýze, které vycházejí z Mahalanobisovy vzdálenosti objektu od centroidů. Odhad pravděpodobnosti zařazení pozorování  $\mathbf{x}$  do skupiny  $k$  je (Tibshirani 2002)

$$\hat{p}_k(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}\delta_k(\mathbf{x})\right)}{\sum_{i=1}^K \exp\left(-\frac{1}{2}\delta_i(\mathbf{x})\right)}.$$

Pokud se sleduje výrazně více znaků než kolik je  $k$  dispozicí objektů, tj.  $p \gg n$ , pak výběrový odhad matice vnitroskupinové variability bude téměř určitě singulární. Nastává tedy problém s nalezením matice, která by vhodně nahradila neexistující inverzní matici k společné vnitroskupinové kovarianční matici, proto jsou odhadnuty pouze její diagonální prvky a ostatní prvky jsou položeny nulové. S rostoucí dimenzí  $p$  při nedostatečném množství pozorování  $n$  je odhad vnitroskupinové kovarianční matice „méně kvalitní“, proto je vhodné zabývat se pouze diagonálními prvky. Nepřesné mimodiagonální prvky mohou zásadním způsobem negativně ovlivnit kvalitu konstruovaného klasifikátoru. Diskriminační skóry metody nejbližších prototypů pak odpovídají restrikcí lineární diskriminační analýzy. Hlavní myšlenka metody nejbližších prototypů je založena na předpokladu, že pro jednotlivé třídy jsou charakteristické jen určité znaky, zatímco ostatní znaky mohou být chápány jako nositelé šumu (Tibshirani 2002; Schlesinger 2004).

### 3.5. Logistická regrese

Regresní analýza se nejčastěji spojuje s regresí lineární, méně často s nelineární nebo logistickou. Cílem regresní analýzy je nalézt co nejlepší, nejúspornější a současně věcně smysluplný model, který popíše vztah mezi závislou proměnnou a skupinou nezávislých proměnných. Je-li vysvětlovaná proměnná spojitá, jedná se o regresi lineární, není-li spojitá, pak o logistickou. Existují metody pro případy, kdy kategorická závislá proměnná je nominální či ordinální – pak se hovoří o multinomické logistické regresi a o ordinální logistické regresi (Řeháková 2000; Pecáková 2007). Logistická regrese byla navržena v 60. letech 20. století (Meloun a Militký 2006).

Binární závislá proměnná  $Y$  nabývá hodnot 0 a 1.  $Y = 1$ , jestliže u sledovaného případu nastal určitý jev; a  $Y = 0$ , jestliže tento jev nenastal. Místo snahy predikovat libovolně zvolené hodnoty sloužící k označení dvou kategorií binární proměnné, se problém zaměřuje na predikci pravděpodobnosti, že případ patří do jedné kategorie závislé proměnné. Známe-li  $P(Y = 1)$ , pak známe i pravděpodobnost jevu opačného, tedy  $P(Y = 0)$ , protože

$$P(Y = 0) = 1 - P(Y = 1).$$

Pravděpodobnost, že  $Y = 1$  lze modelovat jako

$$P(Y = 1) = \alpha + \beta_1 X_1 + \dots + \beta_K X_K.$$

Řešení této regresní rovnice naráží na numerické problémy, protože pravděpodobnost jevu je číslo, které leží mezi nulou a jedničkou, a rovnicí predikované hodnoty by tuto podmínku nemusely splňovat. Prvním krokem k odstranění tohoto nedostatku je záměna pravděpodobnosti jevu šancí jevu. Šance, že  $Y = 1$ , je definovaná jako podíl pravděpodobnosti, že  $Y = 1$  a pravděpodobnosti, že  $Y \neq 1$ , tedy

$$\text{šance}(Y = 1) = \frac{P(Y = 1)}{1 - P(Y = 1)}.$$

Šance nemá žádnou pevnou maximální hodnotu, ale její minimální hodnota je nula. Další transformace spočívá v zlogaritmování šance. Tato proměnná se nazývá logit a je definovaná pomocí vztahu

$$\text{logit}(Y) = \ln \frac{P(Y = 1)}{1 - P(Y = 1)}.$$

Hodnoty logitu se pohybují v intervalu  $(-\infty, +\infty)$ . Použije-li se  $\text{logit}(Y)$  jako závisle proměnná, pak bude mít regresní rovnice tvar

$$\text{logit}(Y) = \alpha + \beta_1 X_1 + \dots + \beta_K X_K.$$

Logit lze převést zpět na šanci pomocí exponenciální funkce jako

$$\text{šance}(Y = 1) = \exp[\text{logit}(Y)] = \exp(\alpha + \beta_1 X_1 + \dots + \beta_K X_K)$$

a od šance zpět k pravděpodobnosti pomocí vztahu

$$P(Y = 1) = \frac{\text{šance}(Y = 1)}{1 + \text{šance}(Y = 1)} = \frac{\exp(\alpha + \beta_1 X_1 + \dots + \beta_K X_K)}{1 + \exp(\alpha + \beta_1 X_1 + \dots + \beta_K X_K)}$$

$$= \frac{1}{1 + \exp(-(\alpha + \beta_1 X_1 + \dots + \beta_K X_K))}$$

Pravděpodobnost, šance a logit jsou tři různé způsoby vyjádření téhož v tom smyslu, že jsou na sebe vzájemně převoditelné. Častou chybou je interpretace hodnot šancí jako pravděpodobností (Řeháková 2000).

### Interpretace regresních koeficientů

Z rovnice  $\text{logit}(Y) = \alpha + \beta_1 X_1 + \dots + \beta_K X_K$  vyplývá, že logistický koeficient  $\beta_k$  lze interpretovat jako změnu logitu spojenou s jednotkovou změnou hodnoty nezávislé proměnné  $X_k$  za předpokladu, že hodnoty ostatních nezávislých proměnných se nezmění. O  $\exp(\beta_k)$  se změní šance, jestliže hodnota nezávislé proměnné  $X_k$  se změní o jednotku a hodnoty ostatních nezávislých proměnných se nezmění. Je-li  $\beta_k > 0$ , šance se zvětší, je-li  $\beta_k < 0$ , šance se zmenší (Řeháková 2000).

### Test významnosti regresních koeficientů

Hodnota regresního koeficientu  $\beta_k$  nestačí k vyslovení závěru, že nezávislá proměnná  $X_k$  je významná pro predikci či vysvětlení závislé proměnné. Test významnosti každého regresního koeficientu je analogický jako u lineární regrese, lze použít  $t$ -test nebo Waldovo testové kritérium (zejména pro velké výběry). Waldova statistika má  $\chi^2$  rozdělení s 1 stupněm volnosti a představuje čtverec poměru odhadu regresního parametru a jeho směrodatné odchylky (Meloun a Militký 2006).

### Volba proměnných

Podobně jako v diskriminační analýze je pro zařazení objektu do skupiny k dispozici zadaný soubor nezávislých proměnných. Základ krokové logistické regresní analýzy tvoří test, zda určitý znak zlepšil kvalitu modelu. Pomocí dopředných či zpětných postupů je určeno, kolik znaků a jaké je vhodné do modelu zařadit (Meloun a Militký 2006).

### Statistiky pro ohodnocení logistického regresního modelu

Statistika  $-2LL$  ( $-2 \log$  likelihood). Tato statistika nabývá kladných hodnot a větší hodnoty indikují horší predikci závislé proměnné. Určí se hodnota této statistiky pro model, který obsahuje jenom konstantu  $\alpha$  a pro model, který obsahuje zvolenou skupinu  $K$  vysvětlujících proměnných. Jejich rozdíl se nazývá věrohodnostní poměr modelu a poskytuje test nulové hypotézy, že v logistickém regresním modelu  $\beta_1 = \beta_2 = \dots = \beta_K = 0$ . Je-li dosažená hodnota statistiky významná, pak je tato hypotéza zamítnuta a zařazení nezávislých proměnných do

modelu umožňuje lepší predikci závislé proměnné, než by byla možná bez této informace (Řeháková 2000).

Pro logistickou regresi bylo navrženo několik analogií ke koeficientu determinace  $R^2$ , např. Cox-Snell  $R^2$ , Nagelkerke  $R^2$ , McFadden  $R^2$ . Interpretace těchto koeficientů je analogická interpretaci koeficientu determinace v lineární regresi (Řeháková 2000). Jakubík (2005) dále uvádí koeficienty Estrella  $R^2$ , dvě varianty Cragg-Uhler  $R^2$  a Veall-Zimmermann  $R^2$ .

Aplikace logistické regrese nevyžaduje žádné předpoklady ohledně apriorních pravděpodobností, tvaru rozdělení nezávislé veličiny či shody kovariančních matic. Náklady chyby prvního a druhého druhu mohou být zohledněny při stanovení klasifikační hranice (jinak 0,5). Logistická regrese je citlivá na multikolinearitu nezávislých proměnných i na výskyt odlehlých pozorování. Ačkoli logistický model nevyžaduje normalitu proměnných, je citlivý i na jejich extrémní nenormalitu (McLeay a Omar 2000).

### 3.6. Probitová regrese

Pokud je pro odhad pravděpodobnosti  $P(Y = 1)$  užitá distribuční funkce normálního normovaného rozdělení, pak

$$P(Y = 1) = \Phi(\alpha + \beta_1 X_1 + \dots + \beta_K X_K),$$

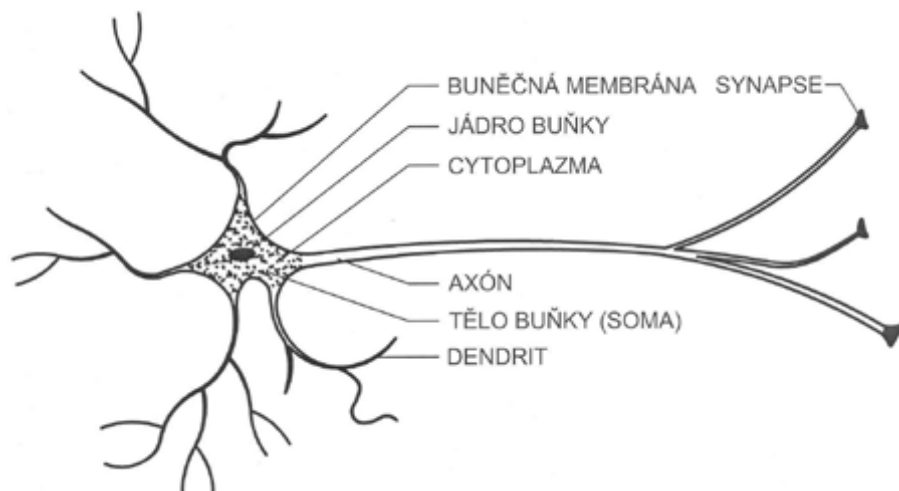
kde  $\Phi()$  je distribuční funkce normálního normovaného rozdělení. Normované normální a normované logistické rozdělení jsou obě symetrická rozdělení s nulovou střední hodnotou, liší se však variabilitou (rozptyl je roven 1, resp.  $\pi^2/3$ ) a špičatostí (3, resp. 4,2). Tyto rozdíly se projevují ve velikosti odhadnutých parametrů; parametry logitového modelu jsou zhruba 1,8 násobkem parametrů probitového modelu. Odhady pravděpodobnosti s užitím logitového a probitového modelu jsou často podobné, logitové modely jsou preferovány – důvodem je zejména jejich snazší interpretace (Hebák et al. 2007).

### 3.7. Neuronové sítě

Za počátek vzniku oboru neuronových sítí je považována práce Warrena McCullocha a Waltera Pittse z roku 1943, kteří vytvořili jednoduchý matematický model neuronu, což je základní buňka nervové soustavy. Číselné hodnoty parametrů v tomto modelu byly z množiny  $\{-1, 0, 1\}$ . Ukázali, že nejjednodušší typy neuronových sítí mohou v principu počítat libovolnou funkci. V roce 1957 Frank Rosenblatt vynalezl perceptron, který je zobecněním McCullochova a Pittsova modelu neuronu pro reálný číselný obor parametrů (Šíma a Neruda 1996).

Základním stavebním funkčním prvkem nervové soustavy je nervová buňka - neuron. Mozková kůra člověka je tvořena zhruba 13 - 15 miliardami neuronů, přičemž každý může být spojen s až 10 tisíci jinými neurony. Neuron je přizpůsoben pro přenos signálů tak, že kromě vlastního těla (soma), má vstupní a výstupní přenosové kanály (dendrity a axon). Z axonu odbočuje řada větví - terminálů, které se převážně stýkají s dendrity jiných neuronů. K přenosu informace slouží mezineuronové rozhraní - synapse (Obrázek 3).

Obrázek 3 Schéma biologického neuronu



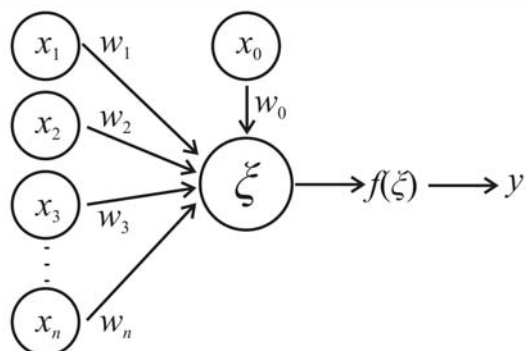
Pramen: Drábek et al. (2005)

### 3.7.1. Matematický model neuronové sítě

#### 3.7.1.1. Formální neuron

Základem matematického modelu neuronové sítě je formální neuron. Jeho struktura je schematicky znázorněna na obrázku (Obrázek 4).

Obrázek 4 Formální neuron



Pramen: Řezanková et al. (2009), vlastní zpracování

### **Vstupy neuronu ( $x_i$ )**

Neuron má  $n$  vstupů  $x_1, x_2, \dots, x_n$ , které modelují dendrity. V závislosti na poloze neuronu v neuronové síti lze vstupy rozdělit do dvou skupin. První skupinu vstupů tvoří podněty z vnějšího okolí, druhou skupinu tvoří vstupy z jiných neuronů.

### **Váhy spojení ( $w_i$ )**

Vstupy neuronu jsou ohodnoceny hodnotou synaptické váhy příslušného spojení,  $w_1, w_2, \dots, w_n$ . Váhy spojení neuronů jsou vyjadřovány reálnými čísly, která vypovídají o průchodnosti či důležitosti spojení. Váhy patří mezi parametry, jejichž změnou během procesu učení je možné dosáhnout shody mezi výstupy zkoumaného procesu a výstupy neuronové sítě. Výpočet vah a jejich ladění představuje podstatnou část učících algoritmů neuronových sítí.

### **Práh neuronu ( $x_0 w_0$ )**

V biologickém neuronu práh představuje hodnotu, kterou vstupní signál neuronu musí překonat, aby se mohl síti šířit dál. Hodnota prahu tedy určuje, kdy bude neuron aktivní či neaktivní. V modelech umělých neuronů je práh používán k posunu signálu při vstupu do aktivační funkce. Práh neuronu se modeluje jako násobek synaptické váhy  $w_0$  a hodnoty fiktivního neuronu  $x_0$ .

### **Vnitřní potenciál neuronu ( $\xi$ )**

Vnitřní (vstupní) potenciál neuronu je vypočten pomocí agregační funkce, jejímž úkolem je sloučit vstupní signály  $x_i$  neuronu. Vnitřní potenciál neuronu  $k$ , který má  $n$  vstupů je potom

$$\xi(k) = \sum_{i=1}^n x_i \cdot w_i + x_0 \cdot w_0, \text{ resp. } \xi(k) = \sum_{i=0}^n x_i \cdot w_i.$$

### **Aktivační funkce ( $f(\xi)$ )**

Aktivační funkce převádí hodnotu vstupního potenciálu na výstupní hodnotu neuronu. Konkrétní tvary aktivačních funkcí bývají různé, obecně je možno je rozdělit na lineární a nelineární. Výběr vhodné aktivační funkce závisí na typu řešené úlohy či na poloze neuronu v neuronové síti (Drábek et al. 2006a). Mezi užívané patří ostrá nelinearita, lineární saturovaná funkce, logistická funkce, hyperbolicko tangenciální funkce, Gaussovská funkce či lineární funkce.

#### **3.7.1.2. Umělá neuronová síť**

Neuronová síť jako celek realizuje určitou transformační funkci, která převádí hodnoty vstupních veličin na hodnoty veličin výstupních. Při modelování a říze-

ní složitých, často nelineárních soustav, vyvstává problém, že daný proces není možné s požadovanou přesností matematicky popsat, nebo je matematický model tak složitý, že jeho algoritmicizace je buď náročná či nemožná. V těchto případech umělé neuronové sítě slouží jako univerzální aproximátor (Drábek et al. 2006b).

### Topologie umělé neuronové sítě

Pojmem topologie (architektura, struktura) neuronové sítě se rozumí konkrétní umístění jednotlivých neuronů a jejich vzájemná propojení. Topologie umělých neuronových sítí je většinou uspořádána do vrstev. Podle polohy vrstvy v síti se pak vrstvy a v nich umístěné neurony dělí na:

- vstupní;
- skryté (pracovní, mezilehlé);
- výstupní.

Vstupní vrstva je tvořena vstupními terminály a slouží ke vstupu signálu z okolí a jeho rozdělení do neuronů následující vrstvy. Tato vrstva není pro její pasivní charakter do celkového počtu vrstev počítána. Poslední, výstupní vrstva slouží k přenosu výstupních signálů z neuronové sítě do okolí. Všechny případné mezilehlé vrstvy se označují jako vrstvy skryté. Jejich úkolem je zvýšit aproximační schopnosti sítě jako celku (Drábek et al. 2006b).

Pro návrh topologie neuronové sítě neexistuje jednoznačné doporučení, obvykle je nutno vycházet ze sérií experimentů. Pro stanovení počtu neuronů ve skryté vrstvě dvouvrstvé sítě  $p$  Taufer et al. (2006c) uvádějí vztahy

$$p = \sqrt{mn} \text{ nebo } p \geq m + n$$

kde  $n$  je počet neuronů ve vstupní vrstvě a  $m$  počet neuronů ve výstupní vrstvě. Pro síť se dvěma skrytými vrstvami

$$p_1 = mr^2, \quad p_2 = mr,$$

kde  $r = \sqrt[3]{n/m}$ .

### Učení neuronové sítě

Proces učení představuje proces, při němž dochází k modifikaci nastavitelných parametrů neuronové sítě za účelem dosažení shody mezi výstupy modelovaného problému a výstupy neuronové sítě. Většinou se proces učení soustředí na adaptaci vah spojení mezi neurony, někdy též např. strmosti aktivačních funkcí či změnu struktury sítě. Metody učení lze rozdělit (obecně, nejen pro neuronové sítě) na učení s učitelem (supervised learning) a učení bez učitele (unsupervised learning). Pro účely práce se budu zabývat pouze metodou učení s učitelem. U těchto sítí se požadované výstupy porovnávají se skutečnými výstupy sítě. Při procesu učení pak dochází k modifikaci synaptických vah (či dalších parametrů) pro dosažení co nejvyšší shody mezi požadovaným a skutečným výstupem. Při procesu učení se tedy hledá minimum chybové (ztrátové)



funkce  $E(w)$ , která je definována jako součet parciálních chyb sítě  $E_k(w)$  vzhledem k jednotlivým tréninkovým vzorům,

$$E(w) = \sum_{k=1}^p E_k(w).$$

Parciální chyba  $E_k(w)$  je vzhledem k  $k$ -tému trénovacímu vzoru úměrná součtu mocnin odchylek skutečných hodnot výstupů sítě pro  $k$ -tý tréninkový vzor a požadovaných hodnot výstupů pro tento vzor,

$$E_k(w) = \frac{1}{2} \sum_{j \in Y} (\hat{y}_j(w, x_k) - y_{jk})^2.$$

Cílem učení je minimalizace chyby sítě  $E(w)$ . Tato chyba závisí na komplikované nelineární složené funkci vícevrstvé sítě, tento cíl tak představuje netriviální optimalizační problém (Šíma a Neruda 1996).

### 3.7.2. Algoritmus učení Back-propagation

Na začátku adaptace v čase 0 jsou váhy  $w$  nastaveny náhodně blízko nuly, např. v intervalu  $\langle -1, 1 \rangle$ . Adaptace probíhá v časových krocích (epochách), které odpovídají tréninkovým cyklům. Nová konfigurace  $w^{(t)}$  v čase  $t > 0$  se vypočte

$$w_{ji}^{(t)} = w_{ji}^{(t-1)} + \Delta w_{ji}^{(t)},$$

kde změna vah  $\Delta w^{(t)}$  v čase  $t$  je úměrná zápornému gradientu chybové funkce  $E(w)$  v bodě  $w^{(t-1)}$

$$\Delta w_{ji}^{(t)} = -\alpha \frac{\partial E}{\partial w_{ji}} w^{(t-1)},$$

kde  $\alpha$  je rychlost učení;  $0 < \alpha < 1$ . Při volbě vhodné rychlosti učení  $\alpha$  metoda vždy konverguje k nějakému lokálnímu minimu. Problémem gradientní metody je, že toto nalezené lokální minimum nemusí být minimem globálním (Šíma a Neruda 1996).

Varianta metody spočívá ve výpočtu hodnot strmostí aktivačních funkcí  $\lambda_j$ , které pak nejsou stanovovány experimentálně. Chyba sítě  $E(w, \lambda)$  je pak kromě synaptických vah také funkcí strmostí. Další modifikace zohledňuje při výpočtu změny váhy ve směru gradientu chybové funkce navíc předchozí změnu vah (tzv. moment setrvačnosti),

$$\Delta w_{ji}^{(t)} = -\alpha \frac{\partial E}{\partial w_{ji}} w^{(t-1)} + \mu \Delta w_{ji}^{(t-1)},$$

kde  $0 < \mu < 1$  je parametr momentu, který určuje míru vlivu předchozí změny. Pomocí momentu gradientní metoda lépe opisuje tvar chybové funkce, protože bere v úvahu předchozí gradient (Šíma a Neruda 1996).

## **Problémy spojené s aplikací neuronových sítí**

1. Volba topologie. Architektura sítě by měla odpovídat složitosti řešeného problému, malá síť nedokáže řešit komplikovaný problém. Bohatá topologie sice může umožnit nalézt globální minimum chybové funkce, ovšem roste výpočetní náročnost adaptace. Nalezená konfigurace může příliš zohledňovat tréninkové vzory včetně jejich chyb a pro nenaucené vzory dávat špatné výsledky (přeučení = přesné zapamatování tréninkové množiny bez zobecnění zákonitostí v ní obsažených).
2. Volba parametrů. Kromě topologie sítě je třeba stanovit řadu dalších parametrů, jako strmosti aktivačních funkcí, rychlost učení, moment setrvačnosti, určit délku trénování.

Taufer et al. (2007) konstatují, že neexistuje paušální doporučení pro výběr topologie, tvaru trénovacích množin a parametrů neuronových sítí a jejich učení. Abid a Zouari (2000) doporučují čtyři základní kritéria pro výběr modelu neuronové sítě: 1) nejlepší klasifikace na trénovacím souboru; 2) nejlepší klasifikace na testovacím souboru; 3) nejmenší rozdíl mezi přesností klasifikace na trénovacím a testovacím souboru a 4) co nejjednodušší architektura neuronové sítě.

### **3.7.3. Genetické učení**

Genetické algoritmy používají operace napodobující evoluční principy živých organismů - selekce, křížení a mutace. Optimalizační metody hledají maximum účelové funkce v závislosti na jejích parametrech. Genetický algoritmus hledá v prostoru všech přípustných řešení ta, která mají maximální hodnotu účelové funkce. Každé řešení v tomto pojetí je označováno jako jedinec, množina jedinců pak populace. Oproti gradientním optimalizačním metodám, které zlepšují jedno řešení, genetický algoritmus udržuje celou populaci řešení. Jednotlivá řešení se navzájem ovlivňují a modifikují pomocí genetických operátorů (Řezanková et al. 2009).

Na počátku je z parametrického prostoru náhodně vygenerována množina jedinců - 1. generace. Pro jednotlivé jedince se spočte funkce charakterizující vhodnost jedince - kriteriální funkce (v případě neuronových sítí se jedná o celkovou chybu sítě). Následující generace je vždy tvořena z předchozí generace (Biskup 2009).

1. Do další generace na základě nejvyšší vhodnosti postupuje část jedinců přímo. Druhá část nově vznikající generace je z předchozí generace vybrána náhodně (selekce).
2. Další část vzniká procesem křížení - zkombinováním genotypů dvou rodičů - výsledkem jsou potomci, u nichž je genotyp vytvořen záměnou některých sobě odpovídajících genů. Záměna genů může být dána nebo je volena náhodně.

3. Poslední skupina nové generace je odvozena od předchozích pomocí mutace - náhodné změny genu.

### **Výhody genetického učení**

Postup při genetickém učení je méně složitý než back-propagation. Při výběru přenosové funkce není nutné omezení na její dvojnásobnou diferencovatelnost nutnou pro gradientní metodu. Křížení a mutace zabraňují uvíznutí algoritmu v lokálních minimech funkce celkové chyby sítě (Biskup 2009).

## **3.8. Klasifikační stromy a klasifikační lesy**

Klasifikační a regresní stromy se pěstují od 60. let 20. století. Silným metodologickým impulsem byla v 80. letech kniha kolektivu Lea Breimana Classification and regression trees, popisující metodu CART (Klaschka a Kotrč 2004).

Model klasifikačního stromu lze popsat stromovým grafem, který se skládá z uzlů a orientovaných hran. V každém neterminálním uzlu se strom větví, z uzlu vedou hrany do dvou či více dceřiných uzlů. Nejběžnější je binární větvení založené na hodnotě jednoho znaku, některé metody umožňují i větvení založená na jejich lineární kombinaci. Pozorování podle hodnot prediktorů postupuje od kořenového uzlu přes větvení v neterminálních uzlech k některému terminálnímu uzlu - listu. Terminálnímu uzlu a zároveň pozorováním, která do něj patří, je přiřazena některá z tříd (Klaschka a Kotrč 2004).

K pěstování stromů prakticky všechny metody využívají tzv. rekurzivní dělení (recursive partitioning). Konstrukce začíná stromem o jediném uzlu, do kterého patří všechna trénovací data. Následně se probere množina všech možných větvení a pro každé z nich se vypočte kritériální statistika, která hodnotí, jak jsou potenciální dceřiné uzly vnitřně homogenní a navzájem odlišné (z hlediska hodnot závisle proměnné). Větvení s maximální hodnotou kritéria se vybere jako nejlepší a použije se v modelu, v němž přibude dvojice uzlů, jež jsou prozatím terminální. Data, která patří do kořenového uzlu, se rozdělí podle hodnot prediktorů mezi nové dceřiné uzly a pro každý z těchto provizorních listů se procedura opakuje (opět se hledá nejlepší větvení) (Klaschka a Kotrč 2004).

Při konstrukci klasifikačního stromu je žádoucí dosáhnout co nejnižší klasifikační chyby, s rostoucí velikostí stromu však klesá (či neroste) chyba na trénovacích datech, ale skutečná chyba často klesá jen do určité úrovně a s dalším zvětšováním stromu opět roste (Klaschka a Kotrč 2004).

Sedlačík (2006) uvádí čtyři základní techniky konstrukce klasifikačního stromu. *Konstrukce zdola nahoru* - v každé úrovni stromu se opakovaně počítají vzdálenosti mezi dvojicemi apriori definovaných tříd a vždy dvě nejbližší jsou spojeny. *Konstrukce shora dolů* - při tomto přístupu je třeba vyřešit tři problémy (výběr dělicího pravidla v každém uzlu; ukončení dělení, tj. rozhodnutí, kdy už je uzel

terminální; přiřazení třídy vysvětlující proměnné každému koncovému uzlu). *Hybridní konstrukce* kombinuje předchozí přístupy. *Konstrukce růst-prořezávání* (growing-pruning). Místo rozhodování, zda je uzel terminální nebo ne, se vytvoří rozsáhlý strom, jehož koncové uzly budou obsahovat pouze data z jedné třídy, a následně se prořeže tak, aby byla minimalizována chyba klasifikace.

Rost a Čermáková (2007) a Rost a Tlustý (2008) popisují tvorbu klasifikačního stromu pomocí algoritmu CART. *První fáze* spočívá ve výběru dělicího kritéria v každém z uzlů pěstovaného stromu. Tento problém je řešen pomocí míry nečistoty  $i(t)$  uzlu  $t$ , přičemž ta je definována tak, aby hodnota této funkce byla co největší, pokud jsou ve zkoumaném uzlu zastoupeny všechny třídy rovnoměrně. Naopak míra nečistoty nabývá nejmenší hodnoty, pokud je ve sledovaném uzlu zastoupena pouze jedna třída. Za nejlepší dělicí kritérium je pak považováno takové kritérium, které způsobí maximální pokles nečistoty v nově tvořeném uzlu. K definování míry nečistoty  $imp(t)$  lze v průběhu růstu klasifikačního stromu použít Giniho index diverzity,

$$imp(t) = \sum_{j \neq k} p(j|t) \cdot p(k|t),$$

kde  $p(j|t)$  představuje pravděpodobnost, že objekt v uzlu  $t$  patří do  $j$ -té skupiny. Tato pravděpodobnost je zpravidla založena na tréninkovém souboru. Jako míru nečistoty uzlu lze kromě Giniho indexu užít i chybu klasifikace nebo křížovou entropii (Hastie et al. 2009). *Druhá fáze* konstrukce klasifikačního stromu spočívá v rozhodnutí, za jakých podmínek se uzel stane listem. K tomuto rozhodnutí se zpravidla používá metoda prořezávání. Princip prořezávání stromu spočívá odvození stromu, jehož koncové uzly jsou buď naprosto čisté nebo počet objektů v koncových uzlech je menší než předem stanovená hodnota. Pomocí odhadu relativní chyby klasifikace je následně odvozen podstrom minimalizující odhad relativní chyby klasifikace. *Třetí fáze* spočívá v přiřazení třídy vysvětlované proměnné každému z koncových uzlů. Tento problém je řešen tak, že přiřadíme koncovému uzlu hodnotu minimalizující odhad chybné klasifikace.

### **Výhody a nevýhody klasifikačních stromů**

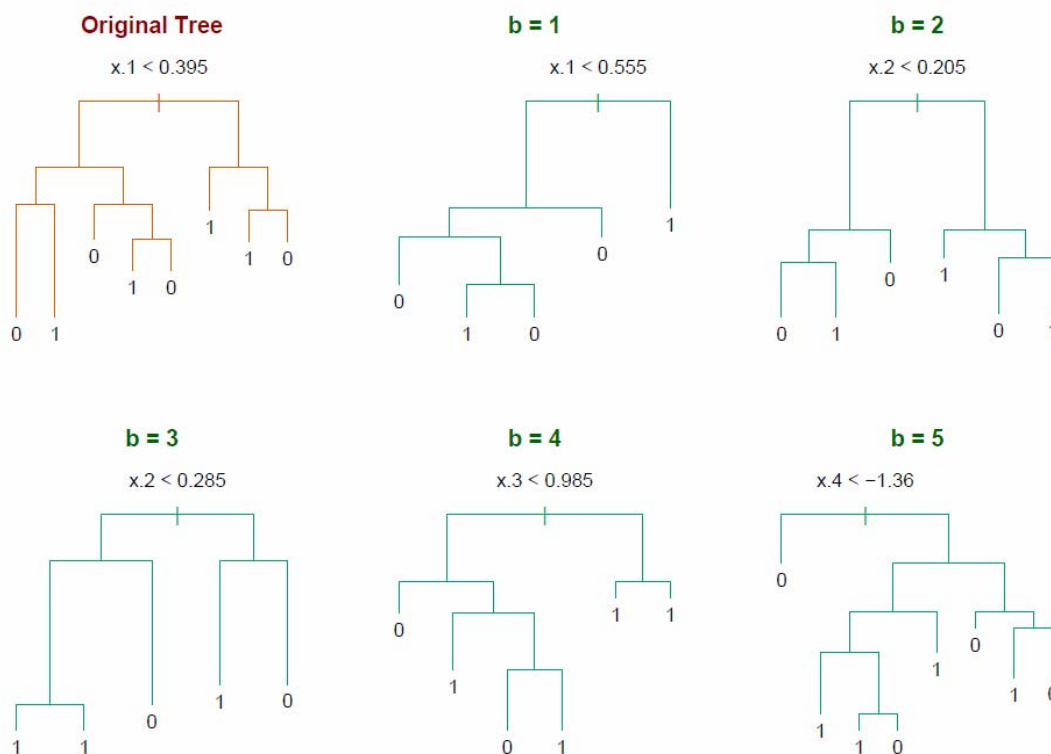
Klasifikační stromy postihují vztahy mezi různými typy veličin, nelineární závislosti, interakce proměnných a závislosti, které mají rozdílnou podobu v různých částech prostoru. Ve srovnání s klasickými parametrickými metodami dosahují často srovnatelné přesnosti, ale poskytují přitom přehlednější a názornější modely. Nevýhodou stromů je, že jsou obvykle značně nestabilní – často pro jeden vstupní soubor existuje mnoho různých stromů s přibližně stejnou chybou a při malé změně dat nebo vstupních parametrů se může výsledný strom výrazně změnit (Klaschka a Kotrč 2004).

### 3.8.1. Klasifikační lesy

Klasifikační les je klasifikační model, jehož klasifikační funkce je dána kombinací (podle vhodně zvoleného pravidla) klasifikačních funkcí určitého počtu klasifikačních stromů. Kombinování klasifikačních funkcí může být složitější, při hlasování může váha hlasu každé klasifikační funkce záviset na chybě stromu na trénovacích datech (přesnější strom má vyšší váhu). Váha hlasu jednotlivého stromu případně nemusí být stejná pro všechny hodnoty vektoru prediktorů  $\mathbf{x}$  - může záviset např. také na tom, jak je list příslušného stromu, do kterého  $\mathbf{x}$  patří, velký (kolik pozorování z trénovacího souboru do něj patří) a jak je čistý (jak výrazná je převaha nejfrekventovanější třídy). Otázkou je, jak na základě trénovacího souboru vypěstovat více stromů. Použije-li se na jedna data opakovaně zvolený algoritmus (např. CART) se stejnými vstupními parametry, je výstupem pokaždé stejný strom. V literatuře jsou uváděny následující postupy, z nichž některé lze aplikovat i na jiné metody než klasifikační stromy (Klaschka a Kotrč 2004).

*Bagging* je zkratka „bootstrap aggregating“. Z trénovacího datového souboru se vytvoří náhodným výběrem s vrácením  $L$  souborů velikosti  $n$  (bootstrapových výběrů) a každý z nich se použije k sestrojení jednoho klasifikačního stromu a výsledný klasifikační les je pak dán většinovým hlasováním se stejnými vahami. Do bootstrapového výběru jsou některá pozorování z  $L$  vybrána opakovaně a některá naopak vůbec (Klaschka a Kotrč 2004). Breiman (1996) uvádí u klasifikačních lesů velikosti  $L = 50$  (les tedy tvoří 50 stromů vypěstovaných pomocí metody CART) v reálných i simulovaných klasifikačních úlohách snížení chyby klasifikace oproti jednomu stromu o 6 až 43 %. Hastie et al. (2009) uvádějí příklad použití baggingu. Na obrázku (Obrázek 5) je původní strom a pět stromů vytvořených na základě bootstrapových výběrů. Na obrázku jsou uvedena sice jen pravidla vrcholových dělení, je však patrné, že jsou využity různé znaky a různé dělicí body.

Obrázek 5 Příklad klasifikačního lesa vytvořeného metodou bagging



Pramen: Hastie et al. (2009), zkráceno

*Boosting* obvykle označuje algoritmus AdaBoost. Předpokládejme klasifikační metodu, která vytváří klasifikační model na základě trénovacích dat a vektoru vah přiřazených jednotlivým pozorováním. Algoritmus konstruuje posloupnost rozdílných modelů s klasifikačními funkcemi tak, že se podle předcházejících výsledků postupně upravují váhy případů. V prvním kroku se použije váhový vektor  $\mathbf{w}_1$ , například rovnoměrné váhy, a vytvoří se model  $T_1$ . V dalších krocích se ke konstrukci modelu  $T_i$  použije váhový vektor  $\mathbf{w}_i$  získaný úpravou vektoru  $\mathbf{w}_{i-1}$  tak, že váhy pozorování chybně klasifikovaných modelem  $T_{i-1}$  se zvýší a váhy pozorování správně klasifikovaných se sníží. Klasifikační metoda se tak stále více soustředí na problematická pozorování, která jsou opakovaně zařazována do nesprávné třídy. Váha modelu při hlasování závisí na chybě modelu na trénovacích datech (Klaschka a Kotrč 2004).

Metoda *Random Forests* je založena na následujících pravidlech (Klaschka a Kotrč 2004).

1. Trénovací soubory pro jednotlivé stromy jsou tvořeny bootstrapovými výběry z vstupního datového souboru.
2. Při volbě větvení pro daný uzel se z prediktorů, které jsou k dispozici, náhodně vybere několik z nich. Poté se nejlepší větvení hledá již klasicky, ale jen mezi těmi větvenými, která jsou založena na vybraných znacích.
3. Pěstují se velké stromy, které se neprořezávají.

Z uvedených metod se Random Forests týká klasifikačních stromů (resp. lesů), ostatní lze aplikovat nejen při konstrukci klasifikačních stromů, ale i na jiné metody.

### **Výhody a nevýhody klasifikačních lesů**

Klasifikační lesy představují zdokonalení metod založených na klasifikačních stromech. Výrazné zpřesnění modelů je dosaženo při mírném zvýšení výpočetní náročnosti. Nevýhodou lesů oproti stromům je ztráta pro stromy charakteristické přehlednosti. Les tvořený desítkami nebo stovkami stromů je podobně jako neuronová síť „černou skříňku“ (Klaschka a Kotrč 2004).

## **3.9. Hodnocení kvality klasifikátorů**

### **3.9.1. Vyhodnocení účinnosti klasifikace**

Pravděpodobnost mylné klasifikace jednotek neznámého původu je významnou informací o kvalitě používaného klasifikátoru. Při odhadování této pravděpodobnosti lze postupovat následujícími způsoby (Hebák et al. 2004).

#### **Resubstituce**

Klasifikátor je použit k třídění těch jednotek, na jejichž základě byl získán. Tento postup však vede k podhodnocení odhadované pravděpodobnosti a pokud klasifikátor nedosahuje dobrých výsledků u jednotek, na jejichž základě byl odvozen, dá se předpokládat, že u jiných jednotek bude fungovat ještě hůře.

#### **Rozdělení souboru**

Další možností pro odhad pravděpodobnosti mylné klasifikace je rozdělit disponibilní datový soubor na dvě části. Na základě jedné části tohoto souboru je odvozen klasifikátor a údaje v jeho druhé části slouží ke zjištění, jak dobře jsou jednotky klasifikovány. Tento postup poskytuje nestranný odhad pravděpodobnosti mylné klasifikace. Nevýhodný je při jeho použití požadavek dostatečně velkého datového souboru, neboť část jeho jednotek je třeba nejprve oddělit. Klasifikátor je kromě toho horší, než by mohl být v případě, že bychom pro jeho odvození použili všechny jednotky v souboru obsažené.

#### **Křížové ověřování**

V tomto případě je datový soubor rozdělen do  $k$  podmnožin o zhruba stejné velikosti. Klasifikátor je odhadnut  $k$ -krát, pokaždé je vynechána jedna

podmnožina, na které je ověřena účinnost klasifikátoru. Zvláštním případem křížového ověřování je tzv. jackknife (kapesní nůž), kdy počet podmnožin je roven rozsahu souboru. Klasifikátor je v tomto případě odhadnut na základě údajů o všech jednotkách v souboru s výjimkou jedné jednotky a následně je zjištěno, zda by byla tato jednotka s užitím kritéria zařazena správně či nikoliv. Tímto způsobem získaný odhad pravděpodobnosti mylné klasifikace je téměř nestranný.

### 3.9.2. Klasifikační matice

Konfrontaci skutečné skupinové příslušnosti jednotek s jejich zařazením na základě zvoleného klasifikačního kritéria obsahuje tzv. klasifikační (též konfusní) matice (Tab. 2), na jejíž diagonále jsou počty jednotek zařazených správně do odpovídající skupiny, ostatní prvky pak představují počet odlišně klasifikovaných jednotek. Podíl správně klasifikovaných jednotek, tedy  $(TP + TN) / n$ , se nazývá hit ratio. Pravděpodobnost mylné klasifikace pak odhadujeme jako podíl chybně klasifikovaných jednotek (Hebák 2004).

Tab. 2 Klasifikační matice

Skutečná třída	Klasifikace jako		Celkem
	1	0	
1	<i>TP</i>	<i>FN</i>	<i>TP + FN</i>
0	<i>FP</i>	<i>TN</i>	<i>FP + TN</i>
Celkem	<i>TP + FP</i>	<i>FN + TN</i>	<i>n</i>

Pramen: Bortlíček (2008)

Symbole použité v tabulce (Tab. 2)

*TP* (True Positives) - pozorování, která jsou ve skutečnosti pozitivní a klasifikační pravidlo je správně zařadilo mezi pozitivní.

*FN* (False Negatives) - pozorování, která jsou ve skutečnosti pozitivní a klasifikační pravidlo je zařadilo mezi negativní.

*FP* (False Positives) - pozorování, která jsou ve skutečnosti negativní a klasifikační pravidlo je zařadilo mezi pozitivní.

*TN* (True Negatives) - pozorování, která jsou ve skutečnosti negativní a klasifikační pravidlo je správně zařadilo mezi negativní.

Statistický test pro diskriminační sílu klasifikační matice v porovnání s modelem pravděpodobnosti vyjadřuje Pressova  $q$ -statistika. Tato míra porovnává počet správných klasifikací s celkovou velikostí výběru a počtem tříd. Vypočtená hodnota se porovnává s kritickou hodnotou  $\chi^2$  rozdělení pro jeden stupeň volnosti při zvolené hladině významnosti  $\alpha$ . Když testové kritérium pře-



kročí kritickou hodnotu, klasifikační matice se jeví statisticky lepší než klasifikační pravděpodobnost. Pressovo  $q$ -testové kritérium je rovno

$$q = \frac{(n - n_s k)^2}{n(k - 1)}$$

kde  $n$  je velikost výběru,  $n_s$  udává počet objektů správně klasifikovaných a  $k$  je počet skupin. Tento test je však citlivý na velikost výběru, velké výběry jsou náchylné snadněji vykazovat statistickou významnost než výběry malé (Meloun a Militký 2004).

### 3.9.3. ROC křivky

ROC křivky se používají pro grafické znázornění a hodnocení klasifikátorů. Byly převzaty z oblasti radiotechniky (ROC - Receiver operating characteristic, pracovní charakteristika přijímače). Tyto křivky dávají do souvislosti senzitivitu a falešnou pozitivitu (Meloun a Militký 2006). Jejich hodnoty udávají četnosti jednotlivých možností a jsou závislé na hodnotě dělicího bodu. Při změně dělicího bodu se změní hodnoty v klasifikační matici (Tab. 2). Senzitivita, specifita, falešná negativita a falešná pozitivita jsou definovány následovně:

**Senzitivita** ( $Se$ ) - relativní četnost správné klasifikace pozitivních případů

$$Se = TP / (TP + FN)$$

**Specifita** ( $Sp$ ) - relativní četnost správné klasifikace negativních případů

$$Sp = TN / (FP + TN)$$

**Falešná negativita** ( $FNr = 1 - Se$ ) - relativní četnost nesprávné klasifikace pozitivních případů

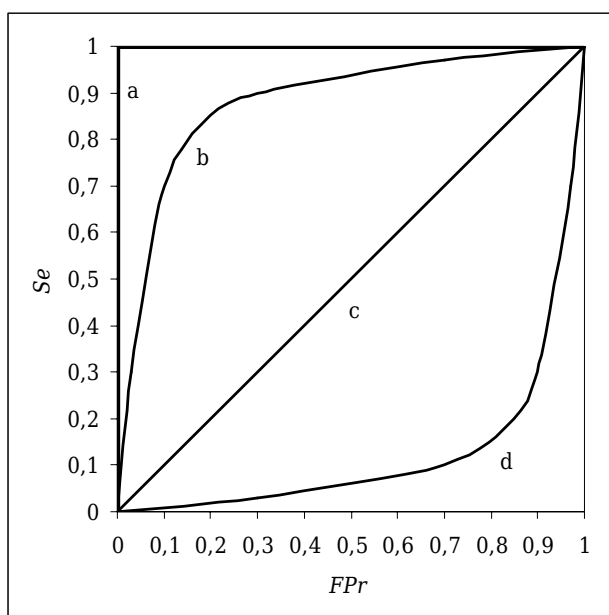
$$FNr = FN / (TP + FN)$$

**Falešná pozitivita** ( $FPr = 1 - Sp$ ) - relativní četnost nesprávné klasifikace negativních případů

$$FPr = FP / (FP + TN)$$

Grafem ROC křivky je dvourozměrný graf, kde na osu  $x$  nanášíme pravděpodobnost špatného zařazení objektů, které jsou ve skutečnosti negativní (falešná pozitivita) a na osu  $y$  pravděpodobnost správného zařazení pozitivních objektů (senzitivita), vzhledem ke všem možným dělicím bodům  $\theta \in \mathbb{R}$  (Obrázek 6). Každému dělicímu bodu  $\theta$  odpovídá právě jeden bod na ROC křivce (Bortlíček 2008).

Obrázek 6 Příklady průběhu ROC křivek

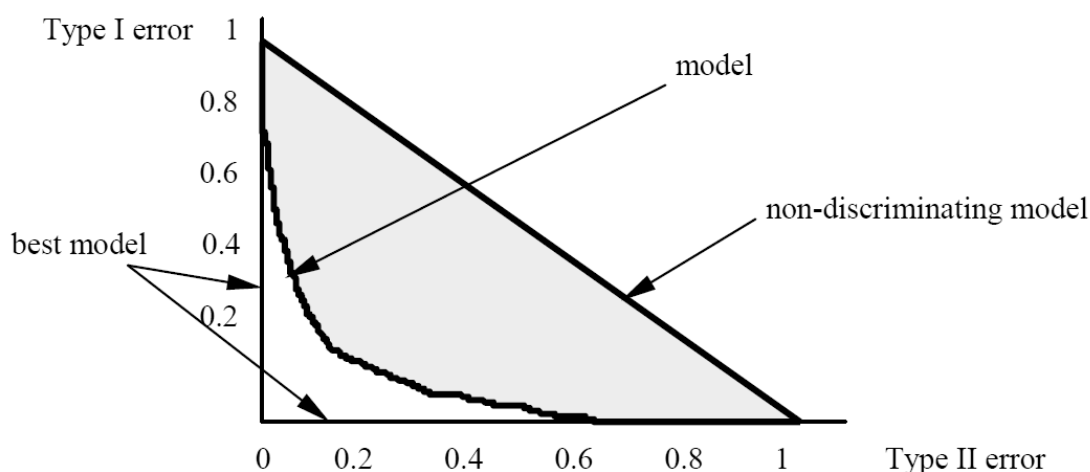


Pramen: Fawcett (2006), vlastní provedení

ROC křivka vždy prochází počátečním bodem  $[0, 0]$  a koncovým bodem  $[1, 1]$ , mezi těmito body je neklesající. Pokud klasifikační pravidlo zařadí všechna pozorování správně (obě chyby budou nulové), výsledkem je ideální ROC křivka, která kopíruje okraj ROC prostoru (z bodu  $[0, 0]$  do bodu  $[0, 1]$  a do bodu  $[1, 1]$ ) – graf křivka a. Tvar ROC křivky závisí na vzájemném překrytí hustot pravděpodobnosti. Pokud jsou hustoty pravděpodobnosti vzájemně silně překryté, pak se ROC křivka přibližuje diagonále. V případě částečného překrytí ROC křivka probíhá mezi diagonálou a ideální ROC křivkou (graf křivka b). Diagonální ROC křivka je extrémní případ, kdy klasifikátor zařazuje objekty náhodně (graf křivka c). Pokud ROC křivka probíhá pod diagonálou, klasifikátor je horší než náhodný prediktor (graf křivka d). V tomto případě lze klasifikátor negovat, takže se z *TP* pozorování stanou *FN* a z *FP* pozorování se stanou *TN* a ROC křivka bude probíhat mezi křivkou náhodného a ideálního klasifikátoru. Klasifikátor tedy nese užitečnou informaci, která byla původně špatně interpretována (Fawcett 2006; Bortlíček 2008).

Ooghe et al. (2005) konstruují obdobný graf, kdy na osu  $x$  vynášejí pravděpodobnost chyby II. druhu a na osu  $y$  pravděpodobnost chyby I. druhu pro všechny možné dělicí body. Získanou křivku nazývají trade-off funkcí, přičemž v případě ideálního klasifikátoru tato funkce splývá s osami (Obrázek 7). Trade-off funkce klasifikátoru, který neurčí správně žádnou jednotku, vyjadřuje funkce  $y = 1 - x$ , tedy součet obou chyb je roven jedné. Vhodnou kvantitativní charakteristikou tohoto grafu je Giniho koeficient. Ten nabývá hodnot z intervalu  $(0; 1)$  a je definován jako poměr plochy mezi trade-off funkcí modelu a nediskriminujícího modelu k ploše mezi ideálním a nediskriminujícím modelem (Ooghe a Spaenjers 2006). Jakubík a Teplý (2007) kumulativní distribuce správných a špatných klasifikací zobrazují pomocí Lorenzovy křivky používané zejména pro znázornění příjmové nerovnosti.

Obrázek 7 Trade-off funkce



Pramen: Ooghe et al. (2005)

### Plocha pod ROC křivkou

Plocha pod ROC křivkou (Area Under the ROC Curve,  $AUC$ ) je kvantitativní index popisující ROC křivku. ROC křivku redukuje na skalární veličinu, kterou je možné použít pro porovnání několika ROC křivek. Pro porovnání lze dále užít i vzdálenost od bodu  $[0, 1]$  (Betinec a Prchal 2006). Plocha pod ROC křivkou je částí jednotkového čtverce, bude tedy vždy nabývat hodnoty mezi 0 a 1. Z popisu průběhu ROC křivek je patrné, že  $AUC$  nebude nabývat menší hodnoty než 0,5, a tedy  $AUC$  bude nabývat hodnot z intervalu  $\langle 0,5; 1 \rangle$  (Bortlíček 2008),

$$AUC = \int_0^1 ROC(p) dp.$$

Fawcett (2008) navrhuje algoritmy pro efektivní generování bodů ROC křivek a pro výpočet plochy pod ROC křivkou pomocí lichoběžníkové metody.

Směrodatná chyba odhadu  $AUC$  je rovna (Hanley a McNeil 1982, Agarwal a Taffler 2007)

$$SE_{AUC} = \sqrt{\frac{AUC(1-AUC) + (n_F - 1)(Q_1 - AUC^2) + (n_{NF} - 1)(Q_2 - AUC^2)}{n_F n_{NF}}},$$

kde  $n_F$  = počet podniků v úpadku (failed),  $n_{NF}$  = počet prosperujících podniků (non-failed),  $Q_1 = AUC / (2 - AUC)$  a  $Q_2 = 2 AUC^2 / (1 + AUC)$ .

Testová statistika  $z$  má normální normované rozdělení,

$$z = \frac{AUC}{SE_{AUC}}.$$

Testová statistika pro porovnání ploch pod ROC křivkami dvou modelů,  $z$ , má normální normované rozdělení,

$$z = \frac{AUC_1 - AUC_2}{\sqrt{SE_{AUC_1}^2 + SE_{AUC_2}^2}}.$$

### 3.9.4. Skutečný podíl chyb

V případě, že klasifikační pravidlo bylo odvozeno na základě lineární diskriminační analýzy (rozdělení náhodného vektoru ve skupinách jsou normální se shodnými kovariančními maticemi) a neuvažují se náklady mylné klasifikace, pak v případě diskriminace do dvou skupin je skutečný podíl chyb (*AER* - Actual error rate) roven (Čermáková 2004)

$$AER = p_1 \Phi \left( -\frac{1}{2} \Delta - \frac{1}{\Delta} \ln \frac{p_1}{1-p_1} \right) + p_2 \Phi \left( -\frac{1}{2} \Delta - \frac{1}{\Delta} \ln \frac{p_1}{1-p_1} \right),$$

v případě vyvážených souborů

$$AER = \Phi \left( -\frac{1}{2} \Delta \right),$$

kde  $p_1$  je podíl pozorování z 1. skupiny,  $\Phi$  je distribuční funkce normálního normovaného rozdělení a  $\Delta$  je výběrová Mahalanobisova vzdálenost mezi skupinovými centroidy.

## 4. Klasifikační modely

V této části bude uveden popis vybraných klasifikačních modelů, které jsou významné metodickým přínosem a nebo svým určením. Altman a Narayanan (2002) uvádějí velmi rozsáhlou rešerši klasifikačních modelů, která zahrnuje 43 prací z 22 států. U každého modelu uvádějí použitou klasifikační metodu, použítá vstupní data, způsob definice skupin a dosažené výsledky. Ve studii není obsažen žádný z českých či slovenských modelů.

Holečková (2008) rozlišuje dvě základní skupiny klasifikačních modelů.

- Bankrotní modely vycházejí ze skutečných údajů a snaží se odpovědět na otázku, zda podnik do určité doby zbankrotuje. Jejich úkolem je poskytnout včasné varování před pravděpodobným úpadkem.
- Bonitní modely jsou založené na teoretických poznatcích, doplněných o empirické poznatky finančních analytiků, a klasifikují podnik podle stupně finančního zdraví. Neumaierová a Neumaier (2008) charakterizují bonitní podnik tak, že vlastník podniku může být spokojen s finanční výkonností podniku, protože podnik tvoří pro svého majitele hodnotu.

Mezi těmito skupinami modelů neexistuje striktně vymezená hranice, obě si kladou za cíl přiřadit podniku jednu číselnou charakteristiku, na jejímž základě je posouzeno finanční zdraví firmy. Rozdíl mezi nimi spočívá především v účelu, ke kterému byly vytvořeny (Růčková 2010).

Ověřováním přenositelnosti klasifikačních modelů mezi rozdílnými ekonomikami (v čase a prostoru) se zabývají Ooghe a Balcaen (2002). Provádějí nové odhady parametrů na vlastním vzorku podniků, čímž odstraňují vliv faktorů specifických pro původní vzorek. Jako možné vysvětlující faktory omezené přenositelnosti modelů uvádějí:

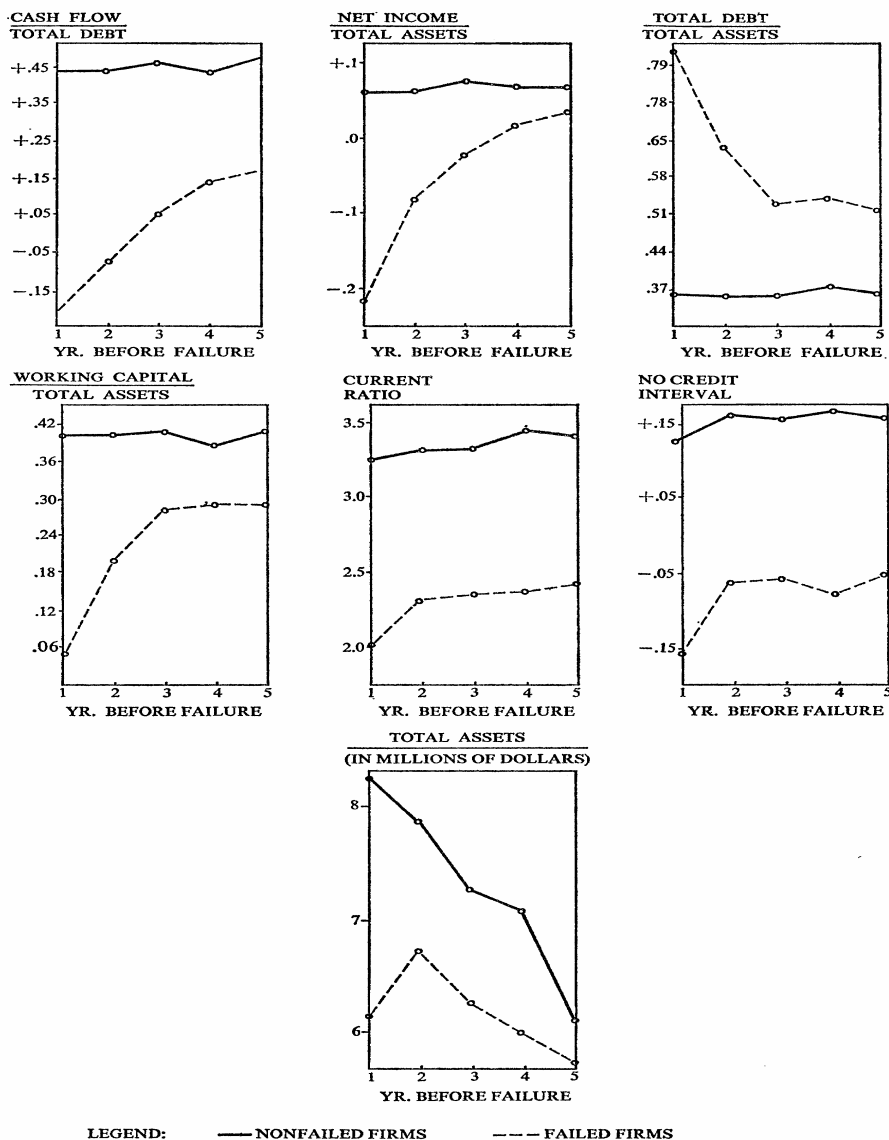
1. Stáří modelu, resp. stáří dat užitých při jeho tvorbě.
2. Země původu modelu, resp. původ podniků, jejichž výkazy byly užity při tvorbě modelu.
3. Vymezení závisle proměnné.
4. Typ podniků, jejichž výkazy byly užity při tvorbě modelu.
5. Použitá klasifikační metoda.
6. Počet, komplexnost a typ proměnných zahrnutých v modelu.

### Beaverova profilová analýza

Beaver (1966) sleduje vývoj třiceti poměrových ukazatelů, které člení do šesti skupin - vybrané ukazatele obsahuje Obrázek 8. Vzorek obsahuje 79 společností v úpadku (podniky, které v letech 1954 - 1964 vyhlásily úpadek, nedodržely závazky plynoucí z emise dluhopisů, přečerpaly bankovní konto či nevyplatily dividendu z prioritních akcií) a 79 prosperujících společností (párování podle odvětví a velikosti). Porovnávání středních hodnot ukazatelů mezi skupinami autor nazývá profilovou analýzou. Dále pro každý ukazatel hledá hraniční bod,

pro který je pravděpodobnost mylné klasifikace minimální. Nejlepších výsledků dosáhl ukazatel Cash flow / Cizí kapitál s chybou 13 % jeden rok před úpadkem, kdy hranicí je hodnota 0,03, resp. 0,07 (pro dva náhodné výběry).

Obrázek 8 Vývoj vybraných poměrových ukazatelů



Pramen: Beaver (1966)

## Index IN95

Index IN95 (Neumaierová a Neumaier 1995) byl sestaven z ukazatelů, které považuje za významné nejvíce modelů finančního zdraví a ve výsledných indikátorech se vyskytují nejčastěji. Váhy ukazatelů indexu IN95 byly stanoveny jako podíl významnosti ukazatele dané četností výskytu daného ukazatele a jeho odvětvové hodnoty v roce 1994. Pro každé odvětví přicházejí v úvahu odlišné váhy jednotlivých ukazatelů (v Tab. 3 pouze zemědělství a ekonomika ČR), pouze u ukazatele *EBIT / U* je váha pro všechny odvětví rovna 0,11 a u ukazatele *OA/KZ* je váha 0,10.

$$IN95 = V_1 \cdot x_1 + 0,11 \cdot x_2 + V_3 \cdot x_3 + V_4 \cdot x_4 + 0,10 \cdot x_5 - V_6 \cdot x_6$$

Znaky:  $x_1$  - aktiva celkem / cizí zdroje  
 $x_2$  - EBIT / nákladové úroky  
 $x_3$  - EBIT / aktiva celkem  
 $x_4$  - výnosy / aktiva celkem  
 $x_5$  - oběžná aktiva / krátkodobé závazky  
 $x_6$  - závazky po lhůtě splatnosti / výnosy

Index IN95 větší než 2 představuje podnik s dobrým finančním zdravím. Podnik s IN95 mezi 1 až 2 není „ani zdraví, ani nemocný“. IN95 menší než 1 znamená podnik finančně neduživý (Neumaierová a Neumaier 1995). Úspěšnost indexu IN95 je více než 70 % (Neumaierová a Neumaier 2002).

Tab. 3 Váhy ukazatelů indexu IN95

Odvětví	Váha ukazatele			
	$V_1$	$V_3$	$V_4$	$V_6$
Zemědělství	0,24	21,35	0,76	14,57
Ekonomika ČR	0,22	8,33	0,52	16,8

Pramen: Neumaierová a Neumaier (1995)

Neumaierová a Neumaier (2008) uvádějí přednosti indexu IN05, které ve značné míře platí i pro jiné modely:

- Jednoduchý výpočet.
- Algoritmy finančních ukazatelů jsou transparentní.
- Využívá veřejně dostupné finanční data o podniku.
- Je možné ho využít pro podniky obchodované i neobchodované na kapitálovém trhu.
- Dává jednoznačné výsledky<sup>4</sup>.
- Je vhodné jej použít jako doplněk zastřešující paralelní ukazatelovou soustavu.

Na druhou stranu uživatelé tohoto indexu musí brát v úvahu:

- Index IN05 byl vytvořen a testován na datech středních a velkých průmyslových podniků.
- Pracuje s ročními daty o výkonnosti podniku.
- Jedná se o orientační charakteristiku. Lze odhadnou celkovou výkonnost podniku, ale neodpovídá na otázku, jak této výkonnosti podnik dosáhl.

## Index IN99

Index IN99 (Neumaierová a Neumaier 2002) se zakládá na datech roku 1999. Pro vzorek 1 698 firem byl vypočten ekonomický zisk. Podniky byly rozděleny

<sup>4</sup> Výsledky nemusejí být vždy jednoznačné, zejména s ohledem na modely obsahující tzv. interval šedé zóny.

na dvě skupiny - firmy s kladnou hodnotou ekonomického zisku a firmy se zápornou hodnotou ekonomického zisku. Pomocí lineární diskriminační analýzy byly zprostředkovány ukazatele nejlépe vysvětlující rozdíl mezi oběma skupinami a jejich významnost odráží hodnota jejich vah. Autoři uvádějí úspěšnost klasifikace 85 %.

Znaky:  $x_1$  - cizí zdroje / aktiva celkem  
 $x_2$  - *EBIT* / aktiva celkem  
 $x_3$  - výnosy / aktiva celkem  
 $x_4$  - oběžná aktiva / krátkodobé závazky

$$IN99 = -0,017 x_1 + 4,573 x_2 + 0,481 x_3 + 0,015 x_4$$

Pravidlo pro klasifikaci

Kladný ekonomický zisk	$IN99 > 2,07$
Šedá zóna	$0,684 \leq IN99 \leq 2,07$
Záporný ekonomický zisk	$IN99 < 0,684$

### Index IN01

Index IN01 byl vytvořen v roce 2002 s cílem spojit oba předchozí indexy. Celkem 1 915 průmyslových podniků bylo rozděleno na podniky tvořící hodnotu (583 podniků), podniky v bankrotu nebo těsně před ním (503 podniků) a ostatní podniky (829 podniků) a pomocí lineární diskriminační analýzy byl stanoven index IN01 (Neumaierová a Neumaier 2002).

Znaky:  $x_1$  - aktiva celkem / cizí zdroje  
 $x_2$  - *EBIT* / nákladové úroky  
 $x_3$  - *EBIT* / aktiva celkem  
 $x_4$  - výnosy / aktiva celkem  
 $x_5$  - oběžná aktiva / krátkodobé závazky

$$IN01 = 0,13 x_1 + 0,04 x_2 + 3,92 x_3 + 0,21 x_4 + 0,09 x_5$$

Pravidlo pro klasifikaci

Podnik tvoří hodnotu	$IN01 > 1,77$
Šedá zóna	$0,75 \leq IN01 \leq 1,77$
Podnik spěje k bankrotu	$IN01 < 0,75$

Spolehlivost klasifikace autoři uvádějí 67 % pro podniky tvořící hodnotu a 86 % pro podniky spějící k bankrotu.

### Index IN05

Neumaierová a Neumaier (2005) aktualizují model IN01 na datech roku 2004, čímž vzniká model IN05. Změna vah je minimální, dochází však ke změně klasifikačního pravidla.

$$IN05 = 0,13 x_1 + 0,04 x_2 + 3,97 x_3 + 0,21 x_4 + 0,09 x_5$$

Význam symbolů odpovídá předchozímu modelu IN01. U tohoto modelu již autoři omezují vliv nestandardizovaného ukazatele *EBIT / U*, a to při úrocích



blíživých se nule maximální hodnotou 9. Úspěšnost klasifikace ověřují zvláště pro malé (75 %), střední (81 %) a velké (80 %) podniky. Pravidlo pro klasifikaci:

Podnik tvoří hodnotu	$IN05 > 1,6$
Šedá zóna	$0,9 \leq IN05 \leq 1,6$
Podnik spěje k bankrotu	$IN05 < 0,9$

### Altmanův index Z68

Altmanův index důvěryhodnosti (Z-score, Z68) vznikl v roce 1968 (Altman 1968). Do empirického materiálu byla zařazena skupina podniků před bankrotem (33 podniků) a skupina excelentních podniků (33 podniků). Pomocí lineární diskriminační analýzy byly zpro středkovány ukazatele nejlépe rozlišující obě skupiny firem a jejich váhy.

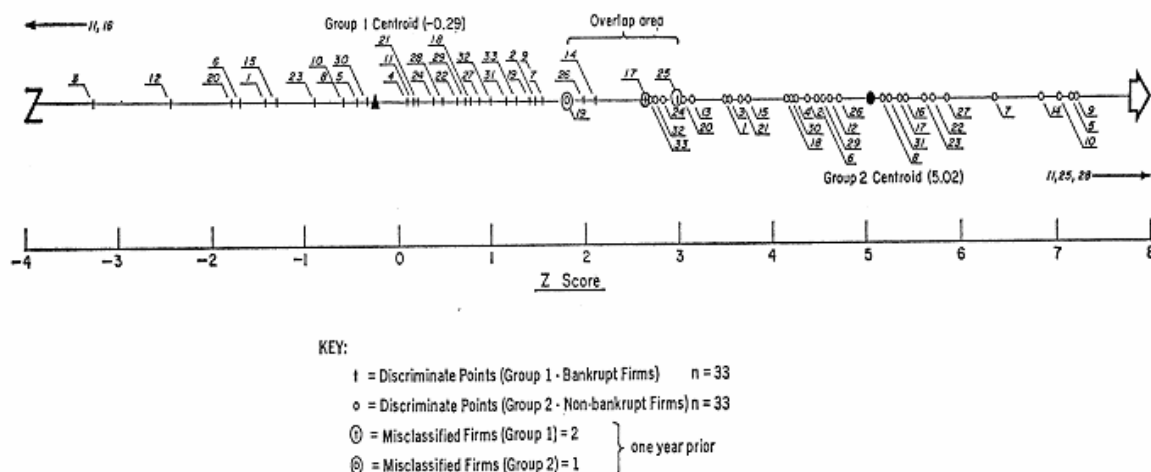
Znaky:  $x_1$  - čistý pracovní kapitál / aktiva celkem (v %)  
 $x_2$  - nerozdělený zisk / aktiva celkem (v %)  
 $x_3$  - EBIT / aktiva celkem (v %)  
 $x_4$  - tržní hodnota vlastního kapitálu / cizí zdroje (v %)  
 $x_5$  - výnosy / aktiva celkem

$$Z68 = 0,012 x_1 + 0,014 x_2 + 0,033 x_3 + 0,006 x_4 + 0,999 x_5$$

Autor uvádí i skupinové centroidy, a to  $-0,29$  u bankrotní skupiny a  $5,02$  u skupiny nebankrotní, z čehož vyplývá, že hranicí pro klasifikaci je hodnota  $2,66$  (Obrázek 9). Na základě mylných klasifikací pak vytvořil interval tzv. šedé zóny  $\langle 1,81; 2,99 \rangle$ . Pravidlo pro klasifikaci:

Úspěšné podniky	$Z68 > 2,99$
Podniky spějící k bankrotu	$Z68 < 1,81$

Obrázek 9 Diskriminační skóre a skupinové centroidy



Pramen: Altman (1968)

Spolehlivost klasifikace ověřená pomocí resubstituce je 95,4 %, přičemž chyba I. druhu činí 6 % a chyba II. druhu 3 %. Spolehlivost klasifikace dva roky před bankrotem je 83 % (28 % chyba I. druhu a 6 % chyba II. druhu).

Altman (1968) patří mezi nejčastěji citované práce v oboru, řada prací používá tento model jako základ pro komparaci výsledků, autoři provádějí testy a odhady parametrů pro rozličné podmínky.

### **Altmanův index $Z'$**

Původní Altmanova diskriminační funkce byla později revidována a aktualizována (Altman 2002). Jindřichovská a Blaha (2001) uvádějí, že na mladých trzích není ještě kursotvorná funkce veřejného obchodování dostatečně vyvinuta a tržní ceny jsou často vzdálené od teoreticky „správné“, opodstatnitelné, rovnovážné ceny, a tudíž jsou zavádějící. U proměnné  $x_4$  byla tržní hodnota vlastního kapitálu u podniků neobchodovaných na kapitálovém trhu nahrazena hodnotou účetní. Proměnnou  $x_4$  tvoří podíl (vlastní kapitál / cizí zdroje), ostatní proměnné zůstávají nezměněny. Proměnná  $x_4$  odvíjející se od účetní a nikoliv od tržní hodnoty umožňuje použít tento model na nekotované firmy, na mladých trzích ale i na rozvinutých kapitálových trzích. Při ověření účinnosti klasifikace bylo 4,5 % objektů chybně klasifikováno.

$$Z' = 0,717 x_1 + 0,847 x_2 + 3,107 x_3 + 0,42 x_4 + 0,998 x_5$$

Pravidlo pro klasifikaci

Úspěšné podniky	$Z' > 2,90$
Podniky spějící k bankrotu	$Z' < 1,23$

### **Altmanův index $Z''$**

Další modifikace určená pro nevýrobní podniky spočívá v odstranění ukazatele (výnosy / aktiva). Tím se minimalizuje potenciální „sektorový efekt“, který je někdy citlivý na velikost odvětví zařazených do testování a může způsobit zkreslení výsledků (Altman 2002). Stejně jako původní rovnice byl i tento model odvozen pomocí lineární diskriminační analýzy na skupinách stejných rozsahů.

Znaky:  $x_1$  - čistý pracovní kapitál / aktiva celkem  
 $x_2$  - nerozdělený zisk / aktiva celkem  
 $x_3$  - EBIT / aktiva celkem  
 $x_4$  - vlastní kapitál / cizí zdroje

$$Z'' = 6,56 x_1 + 3,26 x_2 + 6,72 x_3 + 1,05 x_4$$

Pravidlo pro klasifikaci

Úspěšné podniky	$Z'' > 2,6$
Podniky spějící k bankrotu	$Z'' < 1,1$

### **Altmanův ZETA model**

Altman et al. (1977) odvozuje na základě lineární diskriminační analýzy nový model, ZETA. Jeho cílem je zvýšit přesnost předchozích modelů ve smyslu doby do úpadku. Do vzorku jsou zahrnuty průmyslové i obchodní společnosti, 53 bankrotních a 58 prosperujících. Autor porovnává spolehlivost modelů konstru-

ovaných na základě jak lineární, tak kvadratické diskriminační analýzy, přičemž jejich spolehlivost je jeden rok před bankrotem shodná (92,8 %), v dalších letech klasifikuje lépe LDA, rozdíl je patrný zejména u chyby I. druhu. Vektor koeficientů, kovarianční matice ani skupinové centroidy nejsou z důvodu ochranné známky uvedeny. Autor provádí odhady nákladů mylné klasifikace, a to náklady chyby I. druhu 0,7 a chyby II. druhu 0,02; a apriorních pravděpodobností ( $p_1 = 0,02$  a  $p_2 = 0,98$ ). Z množiny 28 ukazatelů jsou pomocí krokových metod vybrány:

- Zisk před úroky a zdaněním / Celková aktiva
- Stabilita zisku (směrodatná chyba odhadu  $x_1$  pro 5 až 10 roků)
- $\log[\text{Zisk před úroky a zdaněním} / (\text{Úroky} + \text{leasingové platby})]$
- Nerozdělený zisk / Celková aktiva
- Běžná likvidita
- Vlastní kapitál / Celkový kapitál (z pětiletých průměrů)
- $\log(\text{Celková aktiva})$

### **Gurčíkův index**

Gurčík (2002) dělí soubor 60 zemědělských podniků do dvou skupin - prosperující a neprospívající. Za prosperující podnik považuje ten, který v letech 1998 až 2000 dosahoval zisku a v posledním roce rentabilita vlastního kapitálu byla vyšší než 8 %, za neprospívající ostatní podniky. Z výchozích 35 ukazatelů na základě  $t$ -testů shody skupinových středních hodnot a subjektivního posouzení vybírá 5 ukazatelů:

- $x_1$  - nerozdělený hospodářský výsledek / pasiva celkem
- $x_2$  - hospodářský výsledek před zdaněním / pasiva celkem
- $x_3$  - hospodářský výsledek před zdaněním / výnosy
- $x_4$  - cash flow / pasiva celkem
- $x_5$  - zásoby / výnosy

Na tyto ukazatele aplikuje lineární diskriminační analýzu, jejímž výsledkem je diskriminační funkce:

$$G = 3,412 x_1 + 2,226 x_2 + 3,277 x_3 + 3,149 x_4 - 2,063 x_5$$

Pro zatřídění podniků do skupin uvádí pravidlo pro klasifikaci

- |                       |                  |
|-----------------------|------------------|
| prosperující podnik   | $G \geq 1,8$     |
| průměrné podniky      | $-0,6 < G < 1,8$ |
| neprospívající podnik | $G \leq -0,6$    |

Autor uvádí, že index umožňuje rozdělit podniky na prosperující a neprospívající, předpověď bankrotu sám označuje za odvážné tvrzení.

### **Tafflerův model**

Prof. Taffler odvodil několik modelů pro předpověď úpadku podniku, přičemž pravděpodobně nejznámější (často se objevuje v relevantní literatuře) je model z roku 1977. Popis modelu je pak uveden i v autorových dalších pracích, např.

Taffler (1984)<sup>5</sup>. Soubor zahrnoval průmyslové podniky kótované na Londýnské burze cenných papírů a byl rozdělen na skupinu podniků v úpadku (46 podniků) a skupinu úspěšných podniků (46 podniků). Z množiny 80 ukazatelů byly vybrány 4, které k rozlišení skupiny přispívaly nejvíce.

$x_1$  - Zisk před zdaněním / Krátkodobé závazky

$x_2$  - Oběžná aktiva / Cizí kapitál

$x_3$  - Krátkodobé závazky / Aktiva celkem

$x_4$  - No-credit interval<sup>6</sup>

Pomocí lineární diskriminační analýzy byla odvozena diskriminační funkce, přičemž diskriminační koeficienty byly upraveny tak, aby byl jejich součet roven jedné. Při hodnocení autor doporučuje porovnávat výsledek s nulou ( $T < 0$  znamená ohrožení úpadkem).

$$T = 0,53 x_1 + 0,13 x_2 + 0,18 x_3 + 0,16 x_4$$

### **Bodové modely**

Bodový model je jednoduchý systém, který obsahuje několik ukazatelů. Podniku je podle hodnot ukazatelů zahrnutých v modelu přisouzen určitý počet bodů z daného intervalu. Vyšší počet bodů obvykle indikuje lepší finanční situaci. Jednotlivým ukazatelům je přiřazen počet bodů na základě jejich váhy - důležitosti, která může být stanovena podle subjektivního úsudku.

Na základě bodové metody je postaven Tamariho model, Kralickův Quicktest (Kralicek 1993), Metodika výpočtu finančního zdraví pro Operační program Rozvoj venkova a multifunkční zemědělství a Program rozvoje venkova vycházející z Rosochatecká a Řezbová (2004). Mezi tyto jednoduché modely je možné zařadit i např. model navržený v Ooghe et. al (2005), kde je vybraným poměrovým ukazatelům pomocí logitové transformace upraveno měřítko, je jim intuitivně přiřazeno znaménko a jako klasifikátor slouží jejich aritmetický průměr.

### **Kralickův Quicktest**

Kralickův Quicktest (Kralicek 1993) pracuje se čtyřmi ukazateli uvedenými v tabulce (Tab. 4). Každý ukazatel hodnotí pomocí pětibodové škály (Tab. 5), přičemž celková známka se získá jako prostý aritmetický průměr těchto dílčích známek.

---

<sup>5</sup> Původní prací je Taffler, R. J. a Tisshaw H. J. Going, going, gone, four factors which predict. Accountancy, 1977, Vol. 88, No. 1003, pp. 50-52.

<sup>6</sup> Graham (2000) definuje ukazatel No-credit interval jako  $NCI = (\text{Oběžná aktiva} - \text{Zásoby} - \text{Krátkodobé závazky}) / \text{Denní provozní náklady (bez odpisů)}$

Tab. 4 Kralickův Quicktest - ukazatele

Ukazatel	Vzorec
(1) Kvóta vlastního kapitálu	Vlastní kapitál / celkový kapitál • 100
(2) Cash flow v % podnikového výkonu	Cash flow / podnikový výkon • 100
(3) Rentabilita celkového kapitálu	(EBIT + úroky) / aktiva • 100
(4) Doba splácení dluhu v letech	(Cizí kapitál – likvidní prostředky) / roční cash flow

Pramen: Kralicek (1993)

Tab. 5 Kralickův Quicktest - hodnotící tabulka

Ukaza- tel	Výborný (1)	Velmi dobrý (2)	Dobrý (3)	Špatný (4)	Ohrožený insol- ventností (5)
(1)	> 30 %	> 20 %	> 10 %	> 0 %	negativní
(2)	> 10 %	> 8 %	> 5 %	> 0 %	negativní
(3)	> 15 %	> 12 %	> 8 %	> 0 %	negativní
(4)	< 3 roky	< 5 let	< 12 let	> 12 let	> 30 let

Pramen: Kralicek (1993)

### Index bonity

Kromě bodového modelu Kralicek (1993) uvádí i model vytvořený na základě diskriminační analýzy, založený na šesti poměrových ukazatelích. Tento model se v další literatuře (Sedláček 2009) označuje jako Index bonity. Přesné podmínky tvorby modelu (četnosti ve skupinách, popis skupin, výslednou účinnost) však neuvádí.

Znaky:  $x_1$  - Cash-flow / závazky  
 $x_2$  - Bilanční součet / závazky  
 $x_3$  - Výsledek před daní z příjmu / bilanční součet  
 $x_4$  - Výsledek před daní z příjmu / podnikový výkon  
 $x_5$  - Zásoby / podnikový výkon  
 $x_6$  - Podnikový výkon / Bilanční součet

Diskriminační funkce

$$IB = 1,5 x_1 + 0,08 x_2 + 10 x_3 + 5 x_4 + 0,3 x_5 + 0,1 x_6$$

Finančně-ekonomická situace hodnoceného podniku je tím lepší, čím vyšší je hodnota *IB*. Přesnější klasifikaci poskytuje následující stupnice (Sedláček 2009).

extrémně špatná	velmi špatná	špatná	určité problémy	dobrá	velmi dobrá	extrémně dobrá
-2	-1	0	1	2	3	

## Finanční zdraví OP RVMZ a PRV

Rosochatecká a Řezbová (2004) navrhuji bodový model na základě deseti ukazatelů, který byl po úpravě využíván pro hodnocení žadatelů o podporu z Operačního programu Rozvoj venkova a multifunkční zemědělství a z Programu rozvoje venkova. U všech ukazatelů definují návaznost na účetní výkazy a jejich bodové ohodnocení (Tab. 6).

### Původních 10 ukazatelů

1. ROA = (Hospodářský výsledek za účetní období + Nákladové úroky) / Aktiva celkem • 100
2. Dlouhodobá rentabilita = (Fondy ze zisku + Hospodářský výsledek minulých let) / Aktiva celkem • 100
3. Přidaná hodnota/vstupy = Přidaná hodnota / (Náklady vynaložené na prodané zboží + Výkonová spotřeba) • 100
4. Výkonnost přidané hodnoty = (Přidaná hodnota – Osobní náklady – Odpisy dlouhodobého nehmotného a hmotného majetku) / Přidaná hodnota • 100
5. Celková zadluženost = Cizí zdroje / Pasiva celkem • 100
6. Úrokové krytí = (Hospodářský výsledek za účetní období + Nákladové úroky) / Nákladové úroky
7. Doba obratu krátkodobých závazků = (Krátkodobé závazky + Běžné bankovní úvěry + Krátkodobé finanční výpomoci) / (Tržby za prodej zboží + Výkony) • 360
8. Obrátkovost aktiv = (Tržby za prodej zboží + Výkony) / Aktiva celkem
9. Celková likvidita = (Zásoby + Krátkodobé pohledávky + Finanční majetek) / (Krátkodobé závazky + Běžné bankovní úvěry + Krátkodobé finanční výpomoci)
10. Podíl pohledávek a závazků z obchodního styku = Krátkodobé pohledávky z obchodního styku / Krátkodobé závazky z obchodního styku

V konečné verzi, která byla užívána pro hodnocení žadatelů o podporu, byly vynechány ukazatele 7, 8 a 10. U některých ukazatelů došlo k drobným úpravám či byly pozměněny hranice pro hodnocení a bodové hodnocení. Výsledný model vznik doplněním ukazatelů:

1. Doba splatnosti dluhů, z cash flow = (Cizí zdroje – Rezervy – Krátkodobý finanční majetek) / (Výsledek hospodaření za běžnou činnost + Odpisy dlouhodobého nehmotného a hmotného majetku)
2. Krytí zásob ČPK = (Oběžná aktiva + Časové rozlišení – Krátkodobé závazky – Krátkodobé bankovní úvěry – Krátkodobé finanční výpomoci – Časové rozlišení) / Zásoby

Tab. 6 Bodové ohodnocení ukazatelů dle metodiky výpočtu finančního zdraví

Ukazatel	Původní hodnocení			Upravené hodnocení		
ROA (%) MAX	< 1,99	2 až 6	>6,01	< 1,49	1,5 až 3	>3,01
Body	1	2	3	1	2	3
Dlouhodobá rentabilita (%) MAX	< 1,99	2 až 8	>8,01	< 1,99	2 až 8	>8,01
Body	1	2	3	1	2	3
Přidaná hodnota/ vstupy (%) MAX	< 11,99	12 až 30	>30,01	< 14,99	15 až 30	>30,01
Body	1	2	3	1	2	3
Výkonnost přidané hodnoty (%) MAX	< 5,99	6 až 16	>16,01	< 5,99	6 až 15	>15,01
Body	1	2	3	1	2	3
Celková zadluženost (%) MIN	< 46,99	47 až 67	>67,01	< 54,99	55 až 70	>70,01
Body	3	2	1	5	3	1
Úrokové krytí (násobek) MAX	< 1,09	1,1 až 2,5	>2,51	< 1,09	1,1 až 2,1	>2,11
Body	1	2	3	1	2	3
Doba obratu krátkodobých závazků (dny) MIN	< 59,99	60 až 110	>110,01			
Doba splatnosti dluhů (roky) MIN				< 4,99	5 až 7	>7,01
Body	3	2	1	5	3	1
Obrátkovost aktiv (násobek) MAX	< 0,79	0,8 až 1,6	>1,61			
Krytí zásob čistým kapitálem (násobek) MAX				< 0,49	0,5 až 0,7	>0,71
Body	1	2	3	1	2	3
Celková likvidita (násobek) MAX	< 1,49	1,5 až 2,5	>2,51	< 1,49	1,5 až 2,0	>2,01
Body	1	2	3	1	2	3
Pohledávky obchod/závazky obchod (násobek) OPTIM	< 0,99	1,0 až 1,5	>1,51			
Body	1	3	2			

Pramen: Rosochatecká a Řezbová (2004), Metodika výpočtu finančního zdraví

Minimální počet bodů pro konečné hodnocení činí 9, maximální 31 bodů. Výpočet se provádí za tři (popř. dva) roky a výsledný počet bodů se určuje prostým aritmetickým průměrem. Podle počtu celkově dosažených bodů byli žadatelé rozdělení do skupin (Tab. 7), přičemž žadatelé v kategoriích A, B a C splňují podmínku finančního zdraví, žadatelé v kategoriích D a E podmínku finančního zdraví nesplňují.

Tab. 7 Definice kategorií dle metodiky výpočtu finančního zdraví

Kategorie	Počet bodů
A	od 25,01 do 31,00
B	od 17,01 do 25,00
C	od 15,01 do 17,00
D	od 12,51 do 15,00
E	od 9,00 do 12,50

Pramen: Metodika výpočtu finančního zdraví

## Bilanční analýza Rudolfa Douchy

### Bilanční analýza I

Jedná se o pomůcku, která vychází z obvyklých praktik finančních analýz. U všech ukazatelů (jak v jednotlivých oblastech hodnocení, tak i u ukazatele celkového) je stejné vyhodnocení výsledku. Hodnoty větší než 1 jsou dobré, hodnoty mezi 1 a 0,5 jsou považovány za únosné a pod 0,5 za alarmující (Doucha 1996).

Ukazatel stability,  $S = \text{Vlastní kapitál} / \text{Stálá aktiva}$

Ukazatel likvidity,  $L = (\text{Finanční majetek} + \text{Pohledávky}) / (2,17 \cdot \text{Krátkodobý cizí kapitál})$

Ukazatel aktivity,  $A = \text{Celkové výkony} / (2 \cdot \text{Celková pasiva})$

Ukazatel rentability,  $R = 8 \cdot \text{Hospodářský výsledek} / \text{Základní kapitál}$

Celkové hodnocení,  $C = (2S + 4L + A + 5R) / 12$

### Bilanční analýza II

Bilanční analýza II poskytuje soustavu ukazatelů hodnotících podnik ve čtyřech základních okruzích a následně souhrnným ukazatelem. V každém okruhu systém užívá tři až pět ukazatelů, výsledný ukazatel je jejich váženým průměrem. Jednotlivé ukazatele jsou konstruovány tak, že s rostoucí hodnotou ukazují na lepší se stav. Na všech úrovních hodnocení platí hodnocení jako v Bilanční analýze I (Doucha 1995, Doucha 1996).

### Ukazatele stability společnosti

$S_1 = \text{Vlastní kapitál} / \text{Stálá aktiva}$

$S_2 = 2 \cdot \text{Vlastní kapitál} / \text{Celková pasiva}$

$S_3 = \text{Vlastní kapitál} / \text{Cizí zdroje}$

$S_4 = \text{Celková pasiva} / [5 \cdot (\text{Běžné bankovní úvěry} + \text{Krátkodobé finanční výpomoci} + \text{Krátkodobé závazky})]$

$S_5 = \text{Celková aktiva} / (15 \cdot \text{Zásoby})$



Výsledný ukazatel stability  $S$ , resp.  $S_x$ , je vypočítáván jako vážený průměr dílčích ukazatelů stability.

$$S = (2 S_1 + S_2 + S_3 + S_4 + 2 S_5) / 7$$

Pro společnosti s velmi malými nebo nulovými zásobami je definován ukazatel  $S_x$ .

$$S_x = (2 S_1 + S_2 + S_3 + S_4) / 5$$

### **Ukazatele likvidity**

$L_1 = 2 \cdot$  Finanční majetek / (Běžné bankovní úvěry + Krátkodobé finanční výpomoci + Krátkodobé závazky)

$L_2 =$  (Finanční majetek + Pohledávky) / [2,17 • (Běžné bankovní úvěry + Krátkodobé finanční výpomoci + Krátkodobé závazky)]

$L_3 =$  Oběžné prostředky / [2,5 • (Běžné bankovní úvěry + Krátkodobé finanční výpomoci + Krátkodobé závazky)]

$L_4 = 3,33 \cdot$  Pracovní kapitál / Celková pasiva

$L = (5 L_1 + 8 L_2 + 2 L_3 + L_4) / 16$

### **Ukazatele aktivity**

$A_1 = \frac{1}{2}$  Tržby / Celková pasiva

$A_2 = \frac{1}{4}$  Tržby / Vlastní kapitál

$A_3 = \frac{1}{4}$  Přidaná hodnota / Tržby

$A = (A_1 + A_2 + A_3) / 3$

### **Ukazatele rentability**

$R_1 = 10 \cdot$  Hospodářský výsledek za běžné období / Přidaná hodnota

$R_2 = 8 \cdot$  Hospodářský výsledek za běžné období / Základní kapitál

$R_3 = 20 \cdot$  Hospodářský výsledek za běžné období / Celková pasiva

$R_4 = 40 \cdot$  Hospodářský výsledek za běžné období / Tržby

$R_5 = 1,33 \cdot$  Provozní zisk / (Provozní zisk + Finanční zisk + Mimořádný zisk)

$R = (3 R_1 + 7 R_2 + 4 R_3 + 2 R_4 + R_5) / 17$

Celkový výsledný ukazatel je také váženým průměrem,

$C = (2 S + 4 L + A + 5 R) / 12$ , případně

$C_x = (2 S_x + 4 L + A + 5 R) / 12$

### **Bilanční analýza III**

Bilanční analýza III je nadstavbou systému Bilanční analýza II. Její součástí je i jednoduché cash flow, částečně umožňuje sledovat i pohyby peněžních prostředků. Autor doporučuje tento systém používat, pokud má analytik k dispozici čtvrtletní výkazy alespoň dvou po sobě jdoucích roků (Doucha 1996).

## Tamariho model

Tamariho bodový model předvídá finanční situaci pomocí šesti ukazatelů (Sedláček 2009)<sup>7</sup>.

$T_1$  = vlastní kapitál / cizí zdroje

$T_2$  = vývoj zisku s dvěma možnostmi vyjádření:

a) absolutní vyjádření

b) ukazatel ROA

$T_3$  = current ratio (běžná likvidita)

$T_4$  = výrobní spotřeba / průměrný stav nedokončené výroby

$T_5$  = tržby / průměrný stav pohledávek

$T_6$  = výrobní spotřeba / pracovní kapitál

Tab. 8 Tamariho model

Ukazatel	Interval hodnot	Body
$T_1$	0,51 a více	25
	0,41 - 0,50	20
	0,31 - 0,40	15
	0,21 - 0,30	10
	0,11 - 0,20	5
	do 0,10	0
$T_2$	Posledních 5 let kladné a) a b) > $HK$ <sup>8</sup>	25
	Posledních 5 let kladné a) a b) > $Me$	20
	Posledních 5 let kladné a)	15
	b) > $HK$	10
	b) > $Me$	5
	Jinak	0
$T_3$	2,01 a více	20
	1,51 - 2,00	15
	1,11 - 1,50	10
	0,51 - 1,10	5
	Do 0,50	0
$T_4$	$HK$ a více	10
	$Me - HK$	6
	$DK - Me$	3
	$DK$ a méně	0
$T_5$	$HK$ a více	10
	$Me - HK$	6
	$DK - Me$	3
	$DK$ a méně	0

<sup>7</sup> Původní práce TAMARI, M. Financial Ratios as a Means of Forecasting Bankruptcy. Management International Review, 1966, Vol. 6 (4), pp. 15-34.

<sup>8</sup>  $HK$  = horní kvartil,  $Me$  = medián,  $DK$  = dolní kvartil sledovaného souboru

Ukazatel	Interval hodnot	Body
$T_6$	$HK$ a více	10
	$Me - HK$	6
	$DK - Me$	3
	$DK$ a méně	0

Pramen: Sedláček (2009)

Z tabulky (Tab. 8) je patrné, že rozhodující jsou ukazatele  $T_1$  a  $T_2$ . Výsledkem bodování je Tamariho rizikový index, který má maximální hodnotu 100 bodů. Autor svůj rizikový index verifikoval retrospektivně na 130 průmyslových firmách a jejich dosažených výsledcích za léta 1958 až 1960. Pravděpodobnost vzniku insolventnosti je podstatně akutnější ve firmách s nízkou hodnotou Tamariho rizikového indexu než u firem se středně vysokou nebo vysokou hodnotou tohoto indexu. Z podniků, které v roce 1958 dosahovaly vysoké hodnoty indexu, pouze 3 % v roce 1960 ukončily svou činnost, zatímco z podniků s nízkou hodnotou indexu v roce 1958 ukončilo činnost 52 % (Sedláček 2009).

### Grünwaldův index bonity

Grünwaldův index bonity (GIB) je založen na šesti poměrových ukazatelích, které reprezentují rentabilitu, likviditu a finanční stabilitu. Tyto ukazatele jsou vztaženy k určité přijatelné hodnotě (Grünwald 2001). Grünwaldův index bonity je univerzální, nezávislý na příslušnosti podniku k odvětví (Škarpa 2001).

1. Poměrové ukazatele rentability
  - A. Rentabilita vlastního kapitálu
    - $J = \text{zisk po zdanění} / \text{vlastní kapitál (v \%)}$
    - $j = \text{průměrná zdaněná úroková míra z přijatých úvěrů (v \%)}$
  - B. Rentabilita celkového kapitálu
    - $K = \text{zisk před úroky a zdaněním} / \text{aktiva (v \%)}$
    - $k = \text{průměrná úroková míra z přijatých úvěrů (v \%)}$
2. Poměrové ukazatele likvidity
  - A. Provozní pohotová likvidita
    - $L = (\text{krátkodobé pohledávky} + \text{finanční majetek}) / \text{krátkodobé závazky}$
    - $l = \text{raději více než 1, např. 1,2}$
  - B. Krytí zásob pracovním kapitálem
    - $P = (\text{oběžná aktiva} - \text{krátkodobé závazky} - \text{krátkodobé bankovní úvěry}) / \text{zásoby}$
    - $p = \text{méně než 1, např. 0,7}$
3. Poměrové ukazatele finanční stability
  - A. Doba splácení dluhů
    - $S = (\text{cizí zdroje} - \text{rezervy}) / (\text{zisk po zdanění} + \text{odpisy}) \text{ (v letech)}$
    - $s = \text{mnohem déle než 1 rok, např. 3,5 roku}$
  - B. Úrokové krytí
    - $U = \text{zisk před úroky a zdaněním} / \text{úroky}$
    - $u = \text{i značně více než jedenkrát, např. minimálně 2,5 krát}$

$$GIB = \frac{1}{6} \left( \frac{J}{j} + \frac{K}{k} + \frac{L}{l} + \frac{P}{p} + \frac{S}{s} + \frac{U}{u} \right)$$

S výpočtem Grünwaldova indexu bonity je spjata několik podmínek. Bodové hodnocení každého ukazatele je limitováno maximálním počtem 3 bodů. Minimální hodnocení je limitováno 0 body, případný záporný výsledek se nahradí nulou.

Hodnocení pomocí *GIB* navazuje na pásma finančního zdraví.

- A - pevné zdraví ... *GIB* = 2 body a více a přitom všechny poměrové ukazatele alespoň 1,0 bodu
- B - dobré zdraví ... *GIB* = 1 až 2 body a přitom provozní pohotová likvidita a úrokové krytí alespoň 1,0 bodu
- C - slabší zdraví ... *GIB* = 0,5 až 1 bod a přitom provozní pohotová likvidita alespoň 1 bod
- D - churavění ... *GIB* = méně než 0,5 bodu

### Ohlsonův model

Ohlson (1980) pomocí logistické regrese vytvořil tři modely; 1) pro předpověď jeden rok před bankrotem; 2) dva roky před bankrotem; a 3) rok či dva před bankrotem. Výběrový soubor zahrnoval průmyslové podniky, a to 105 bankrotních a 2 058 prosperujících v letech 1970 až 1976. Do modelu vstupují proměnné:

- $x_1 = \ln(\text{Celková aktiva} / \text{Index HDP})$ , základem indexu je rok 1968;
- $x_2 = \text{Celkové závazky} / \text{Celková aktiva}$ ;
- $x_3 = \text{Čistý pracovní kapitál} / \text{Celková aktiva}$ ;
- $x_4 = \text{Krátkodobé závazky} / \text{Oběžný majetek}$ ;
- $x_5 = 1$  pokud jsou Celkové závazky > Celková aktiva, jinak 0;
- $x_6 = \text{Zisk po zdanění} / \text{Celková aktiva}$ ;
- $x_7 = \text{Vyplacené dividendy} / \text{Celkové závazky}$
- $x_8 = 1$ , pokud Zisk po zdanění je v posledních dvou letech negativní, jinak 0;
- $x_9 = (NI_t - NI_{t-1}) / (|NI_t| + |NI_{t-1}|)$ , kde *NI* je zisk po zdanění.

Tab. 9 Regresní koeficienty jednotlivých modelů

	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$	$b_9$	$b_0$
Model 1	-0,407	6,03	-1,43	0,0757	-2,37	-1,83	0,285	-1,72	-0,521	-1,32
Model 2	-0,519	4,76	-1,71	-0,297	-2,74	-2,18	-0,78	-1,98	0,4218	1,84
Model 3	-0,478	5,29	-0,99	0,062	-4,62	-2,25	-0,521	-1,91	0,212	1,13

Pramen: Ohlson (1980)

Regresní koeficienty modelů uvádí tabulka (Tab. 9). U proměnných  $x_4$ ,  $x_7$  a  $x_9$  je mezi modely patrná nestabilita znamének. Jednotlivé modely klasifikují správně 96,1 % (model 1), 95,6 % (model 2) a 92,8 % (model 3) pozorování.

## Zmijewskiho model

Zmijewski (1984) vytváří probitové modely pravděpodobnosti úpadku, pro každý rok separátní model. Dále se věnuje souvislostem mezi vyvážeností vzorků a přesností klasifikace. Pro rok 1978 je pravděpodobnost úpadku rovna

$$P(B = 1) = \Phi(-4,188 - 2,894x_1 + 4,144x_2 + 0,003x_3).$$

$x_1$  = Čistý zisk / Celková aktiva;

$x_2$  = Celkové závazky / Celková aktiva;

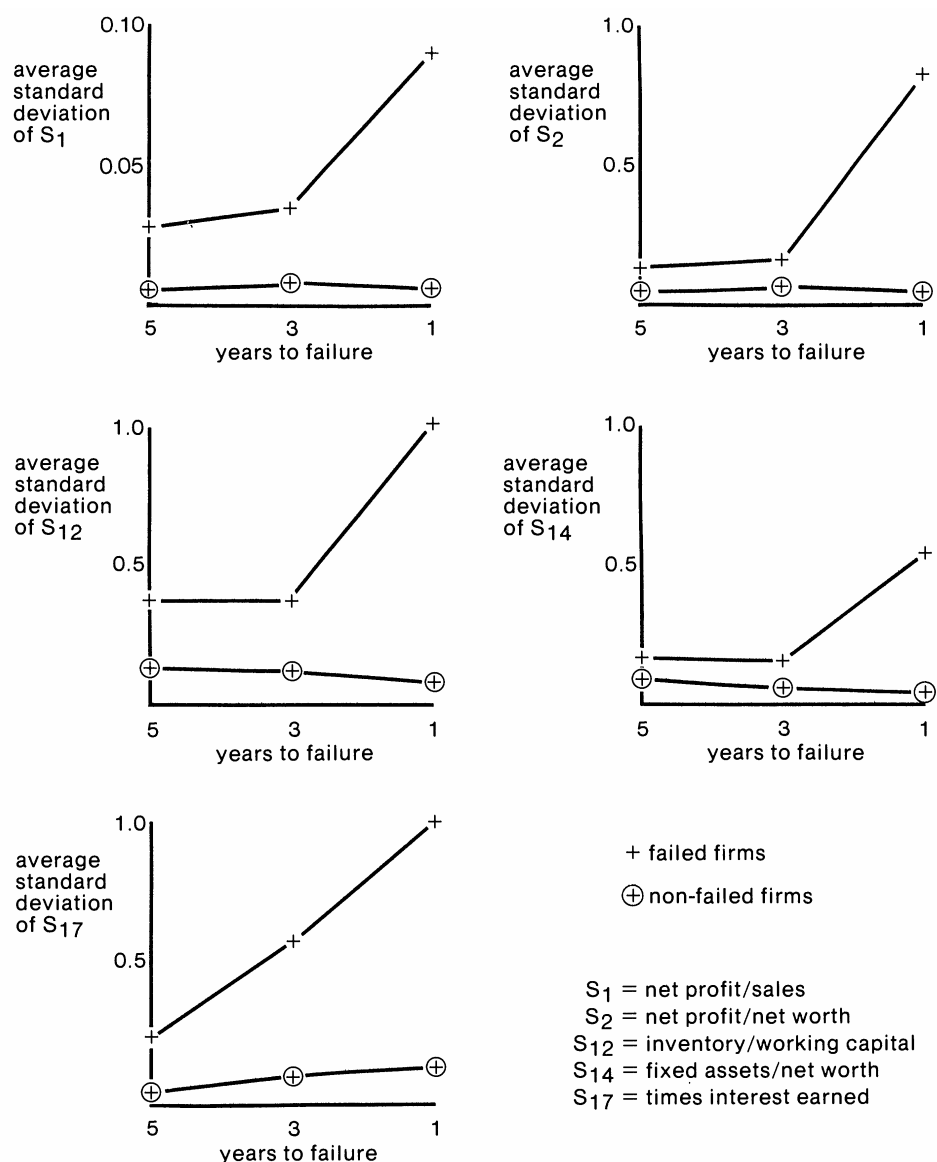
$x_3$  = Oběžný majetek / Krátkodobé závazky.

Celková přesnost klasifikace u tohoto modelu je 99,1 %, prosperujících 99,8 % a bankrotních 29,4 % (vysoká pravděpodobnost chyby I. druhu). Regresní koeficienty jednotlivých modelů jsou poměrně konsistentní, pouze u proměnné  $x_3$  (likvidita) je nestabilní znaménko. V případě uvedeného modelu tedy vyšší likvidita zvyšuje pravděpodobnost úpadku.

## Model obsahující variabilitu

Dambolena a Khoury (1980) nepracují pouze s poměrovými ukazateli, ale vysvětlující proměnné rozšiřují i o ukazatele variability - směrodatné odchylky poměrových ukazatelů, vypočítané za tři (čtyři) po sobě následující období. Ačkoliv vysoká variabilita nemusí nutně znamenat nepříznivý vývoj, empiricky dokazují, že s blížícím se úpadkem se variabilita ukazatelů zvyšuje (Obrázek 10). Základem pro krokovou diskriminační analýzu je 19 poměrových ukazatelů a jejich 19 směrodatných odchylek. Vzorek 46 podniků se skládá z poloviny prosperujících a poloviny upadajících podniků, pro něž byla k dispozici data za 8 let. Autoři srovnávají účinnost klasifikace na základě pouze poměrových ukazatelů a poměrových ukazatelů s ukazateli variability; pro období 1, 3 a 5 let před úpadkem byly vytvořeny samostatné diskriminační funkce. Účinnost klasifikace je při použití ukazatelů variability vyšší, přičemž 1 rok před úpadkem je 95,7 % proti 94,4 %; 3 roky před úpadkem 89,1 % proti 79,7 % a 5 let před úpadkem 82,6 % proti 70,3 %.

Obrázek 10 Průměrné směrodatné odchylky vs. doba před úpadkem



S<sub>i</sub> = average standard deviation of ratio i. Ratio numbers are given in Table 1.

Pramen: Dambolena a Khoury (1980)

### Další klasifikační modely

Abidali a Harris (1995) odvozují predikční model pro odvětví stavebnictví. Z 31 ukazatelů, které zahrnují 24 poměrových ukazatelů a 7 ukazatelů trendu, pomocí lineární diskriminační analýzy sestavují diskriminační funkci se sedmi proměnnými. Vzorek pro aplikaci diskriminační analýzy se skládal z 20 prosperujících podniků a z 11 podniků v úpadku, v obou případech se jednalo o středně velké podniky s více než 50 zaměstnanci.

$$Z = 14,6 + 82 x_1 - 14,5 x_2 + 2,5 x_3 - 1,2 x_4 + 3,55 x_5 - 3,55 x_6 - 3 x_7$$

Znaky:  $x_1$  - Zisk po zdanění / (Stálá aktiva + Čistý pracovní kapitál)

- $x_2$  - Oběžná aktiva / Čistá aktiva<sup>9</sup>
- $x_3$  - Tržby / Čistá aktiva
- $x_4$  - Krátkodobé úvěry / Zisk před zdaněním a úroky
- $x_5$  - Trend daní
- $x_6$  - Trend zisku po zdanění
- $x_7$  - Trend krátkodobých úvěrů

Pro hodnoty trendu ( $T_n$ ) uvádí vztah

$$T_n = \frac{\frac{P_n + P_{n-1}}{2} - P_{n-2}}{|P_{n-2}|},$$

kde  $P$  je sledovaná veličina v letech  $n$ ,  $n - 1$  a  $n - 2$ .

Pravidlo pro klasifikaci

Prosperující podnik	$Z > 2,96$
Šedá zóna	
Podnik v úpadku	$Z < -2,96$

Spolehlivost klasifikace ověřená pomocí resubstituce je 93,5 %, při použití odlišné testovací množiny klesá na 71,6 %.

Lacher et al. (1995) srovnávají Almanův model Z68 s vlastním modelem neuro-nové sítě. Výchozí soubor 94 bankrotních a 188 prosperujících podniků je rozdělen na dvě poloviny - trénovací a testovací množinu. Pro tvorbu sítě používají algoritmus CasCor, který adaptuje nejen synaptické váhy, ale i architekturu sítě. V prvním kroku algoritmu je síť jednovrstvá a v dalších krocích se adaptují váhy a případně přidávají skryté neurony. Spolehlivost klasifikace vytvořené sítě porovnávají s modelem Z68, přičemž celková chyba sítě je 6,4 % v roce bankrotu a 15,9 % tři roky před bankrotem. Celková chyba modelu Z68 je 10,6 % v roce bankrotu a 27,7 % tři roky před bankrotem, přičemž chyba II. druhu je výrazně nižší než chyba I. druhu.

Deakin (1972) se zaměřuje na srovnání diskriminační síly 1 až 5 let před bankrotem. Na základě souboru  $2 \times 32$  podniků a 14 poměrových ukazatelů vytváří 5 diskriminačních funkcí. Spolehlivost klasifikace těchto diskriminačních funkcí je vyšší než u srovnávaných (Beaverův ukazatel cash flow / cizí zdroje a Z68). Tento výsledek lze předpokládat, protože Altmanův model Z68 byl vytvořen na základě poměrových ukazatelů v roce před bankrotem.

Grice a Dugan (2001) na příkladu Ohlsonova a Zmijewkého modelu ověřují stabilitu vztahů mezi finančními ukazateli a finanční tísní. Spolehlivost modelů v odlišných podmínkách, než za jakých byly vytvořeny, je nízká (zejména u Ohlsonova modelu), což omezuje i jejich přenositelnost (do jiných odvětví, do jiné ekonomiky - tedy i v čase).

---

<sup>9</sup> Čistá aktiva = Celková aktiva – Krátkodobé závazky (Abidali a Harris 1995)

## 5. Metodika a materiál

### 5.1. Vymezení podniků ohrožených finanční tísň

Vymezení podmínek, za kterých je podnik označen jako ohrožený finanční tísň, bylo odvozeno z prací Gurčik (2002), Abou El Sood (2008) a Kopta (2009). Gurčik (2002) za prosperující podniky považoval takové, které během tří po sobě následujících roků dosahovaly zisku a v posledním roce rentabilita vlastního kapitálu byla vyšší než 8 % (stanoveno podle míry inflace), za neprosperující ostatní podniky. Abou El Sood (2008) řadil společnost mezi společnosti v úpadku, pokud splňovala během tří po sobě následujících roků jednu z podmínek: 1) záporné provozní cash flow v kterémkoliv roce; 2) čistá provozní ztráta v kterémkoliv roce; nebo 3) záporný čistý pracovní kapitál v kterémkoliv roce. Kopta (2009) vymežil problémové zemědělské podniky jako takové, které byly ohroženy buď dlouhodobou zápornou rentabilitou (součet hospodářských výsledků z běžné činnosti za pět let byl záporný), nebo prudkými výkyvy hospodářského výsledku, které vedly k zápornému provoznímu cash flow. Tato kritéria jsou obzvlášť relevantní, protože autor také analyzoval podniky v odvětví zemědělství, a také dokládal vazbu mezi těmito kritérii a vstupem podniku do konkurzního řízení.

**Podnik je ohrožen finanční tísň, pokud**

- a) součet hospodářských výsledků za tři roky je záporný nebo
- b) pokud je cash flow v kterémkoliv ze tří roků záporné.

Hospodářský výsledek je uvažován jako hospodářský výsledek z běžné činnosti, cash flow na úrovni cash flow ze samofinancování, které je z výkazu zisku a ztráty dopočteno jako

hospodářský výsledek z běžné činnosti

- + odpisy dlouhodobého nehmotného a hmotného majetku;
- + změna stavu rezerv a opravných položek v provozní oblasti a komplexních nákladů příštích období;
- + změna stavu rezerv a opravných položek ve finanční oblasti;
- zisk z prodeje dlouhodobého majetku.

Výkaz zisku a ztráty do roku 2002 nerozděloval tržby a zůstatkovou cenu prodaného majetku na dlouhodobý majetek a materiál, vychází se proto z položek tržby a zůstatková cena prodaného investičního majetku a materiálu.



## 5.2. Výběr ukazatelů

Ukazatele, jejichž schopnost předpovídat finanční tíseň bude ověřována, byly vybrány z široké skupiny ukazatelů uváděných v učebnicích finanční analýzy a používaných v klasifikačních modelech (kapitola 4). Ukazatele jsou vybrány ze všech základních skupin (kromě ukazatelů tržní hodnoty) s důrazem na jejich standardizaci. Je možné je rozdělit do několika skupin, a to na ukazatele

- rentability;
- účinnosti a rychlosti obratu;
- finanční stability;
- nákladovosti;
- založené na peněžních tocích;
- vztahované na počet pracovníků.

Dále bude proveden výčet ukazatelů obsažených v jednotlivých skupinách – jejich název, symbol používaný dále v aplikační části a vztah pro jejich výpočet. Vzhledem k tomu, že se jedná o běžné poměrové ukazatele, nebude uváděn jejich podrobný popis.

### Ukazatele rentability

- míra zisku,  $ZD/A$  (výsledek hospodaření před zdaněním / aktiva)
- rentabilita kapitálu,  $ZUD/A$  [(výsledek hospodaření před zdaněním + nákladové úroky) / aktiva]
- dlouhodobá rentabilita,  $DR$  [(rezervní fondy, nedělitelný fond a ostatní fondy ze zisku + výsledek hospodaření minulých let + výsledek hospodaření běžného účetního období) / aktiva]
- výnosnost,  $ZD/V$  (výsledek hospodaření před zdaněním / výnosy)
- rentabilita výkonů,  $ZD/W$  (výsledek hospodaření před zdaněním / výkony)

### Ukazatele účinnosti a rychlosti obratu, resp. vázanosti

- účinnost aktiv,  $V/A$  (výnosy / aktiva)
- účinnost dlouhodobého majetku,  $V/DM$  (výnosy / dlouhodobý majetek)

U krátkodobého majetku a závazků jsou z důvodu standardizace ukazatelů namísto ukazatelů rychlosti obratu užity ukazatele vázanosti.

- vázanost oběžných aktiv,  $OA/V$  (oběžná aktiva / výnosy)
- vázanost zásob,  $Zas/V$  (zásoby / výnosy)
- vázanost krátkodobých závazků,  $KrZ/V$  (krátkodobé závazky / výnosy)

### **Ukazatele finanční stability**

- zadluženost,  $CZ/A$  (cizí zdroje / aktiva)
- dlouhodobá zadluženost,  $DlCZ/A$  (dlouhodobé cizí zdroje / aktiva)
- krátkodobá zadluženost,  $KrCZ/A$  (krátkodobé cizí zdroje / aktiva)
- krytí aktiv dlouhodobým kapitálem,  $DlK/A$  [(vlastní kapitál + dlouhodobý cizí kapitál) / aktiva]
- krytí dlouhodobého majetku dlouhodobým kapitálem,  $DlK/DM$  [(vlastní kapitál + dlouhodobý cizí kapitál) / dlouhodobý majetek]
- krytí dlouhodobého majetku vlastním kapitálem,  $VK/DM$  (vlastní kapitál / dlouhodobý majetek)
- čistý provozní kapitál na aktiva,  $ČPK/A$  [(oběžná aktiva – dlouhodobé pohledávky – krátkodobé závazky – krátkodobé bankovní úvěry a finanční výpomoci) / aktiva]
- čistý provozní kapitál na výnosy,  $ČPK/V$  [(oběžná aktiva – dlouhodobé pohledávky – krátkodobé závazky – krátkodobé bankovní úvěry a finanční výpomoci) / výnosy]
- běžná likvidita,  $L3$  [oběžná aktiva / (krátkodobé závazky + krátkodobé bankovní úvěry a finanční výpomoci)]
- pohotová likvidita,  $L2$  [(krátkodobé pohledávky + finanční majetek) / (krátkodobé závazky + krátkodobé bankovní úvěry a finanční výpomoci)]
- peněžní likvidita,  $L1$  (finanční majetek / (krátkodobé závazky + krátkodobé bankovní úvěry a finanční výpomoci)]

### **Ukazatele nákladovosti**

- mzdová nákladovost,  $Mz/V$  (osobní náklady / výnosy)
- nákladovost výkonové spotřeby,  $VS/V$  (výkonová spotřeba / výnosy)
- materiálová nákladovost,  $Mt/V$  (spotřeba materiálu a energie / výnosy)
- nákladovost odpisů,  $Odp/V$  (odpisy dlouhodobého nehmotného a hmotného majetku / výnosy)

### **Ukazatele založené na peněžních tocích**

- cash rentabilita aktiv,  $CF/A$  (cash flow ze samofinancování / aktiva)
- cash rentabilita výnosů,  $CF/V$  (cash flow ze samofinancování / výnosy)
- cash rentabilita cizího kapitálu (stupeň oddlužení),  $CF/CZ$  (cash flow ze samofinancování / cizí zdroje)

### **Ukazatele vztahené na počet pracovníků**

- produktivita práce (z výnosů),  $V/PEP$  (výnosy / průměrný evidenční počet pracovníků)

- produktivita práce (z hrubé přidané hodnoty),  $HPH/PEP$  (přidaná hodnota / průměrný evidenční počet pracovníků)
- produktivita práce (z čisté přidané hodnoty),  $ČPH/PEP$  [(přidaná hodnota – odpisy) / průměrný evidenční počet pracovníků]
- hospodářský výsledek na pracovníka,  $ZD/PEP$  (výsledek hospodaření před zdaněním / průměrný evidenční počet pracovníků)

Takto definované ukazatele produktivity práce reflektují změny cenové úrovně pouze v čitateli. Proto je vhodné zařadit peněžní ukazatel i do jmenovatele a definovat produktivitu vztahenou na korunu vynaložených osobních nákladů.

- produktivita práce (z výnosů),  $V/Mz$  (výnosy / osobní náklady) – ukazatel je reciprokou hodnotou ukazatele mzdové nákladovosti
- produktivita práce (z hrubé přidané hodnoty),  $HPH/Mz$  (přidaná hodnota / osobní náklady)
- produktivita práce (z čisté přidané hodnoty),  $ČPH/Mz$  [(přidaná hodnota – odpisy) / průměrný evidenční počet pracovníků]
- hospodářský výsledek na pracovníka,  $ZD/Mz$  (výsledek hospodaření před zdaněním / osobní náklady)

### Dynamika ukazatelů

Hodnocena bude nejen úroveň ukazatelů v daných obdobích, ale rovněž jejich dynamika, a to pomocí „modifikované“ relativní změny. Pro ukazatel  $x$  je tato změna definována jako

$$\delta' = \operatorname{sgn} x_0 \cdot \frac{x_1 - x_0}{x_0} = \frac{x_1 - x_0}{|x_0|} \text{ pro } x_0 \neq 0.$$

Pokud se hodnota  $x$  zvyšuje, pak  $\delta'$  nabývá kladné hodnoty, pokud se hodnota  $x$  snižuje, pak  $\delta'$  nabývá záporné hodnoty bez ohledu na hodnotu  $x_0$  (oproti „klasické“ relativní změně).

## 5.3. Dílčí cíle

Cílem disertační práce je ověření možnosti vícerozměrných klasifikačních metod při predikci finanční tísně podniků v odvětví zemědělství. Dosažení cíle práce bude sledovat následující kroky – dílčí cíle.

1. Ověřit, zda daný problém – tedy předpověď finanční tísně – dokáží řešit klasifikační modely uvedené v části Klasifikační modely. Tedy aplikovat vybrané modely na uvedený materiál a ověřit jejich spolehlivost klasifikace.
2. Posoudit schopnosti jednotlivých ukazatelů a jejich kombinací z hlediska spolehlivosti klasifikace (profilová analýza).

3. Na základě výsledků předchozích bodů ověřit klasifikační schopnosti modelů vytvořených na základě vícerozměrných metod uvedených v kapitole Klasifikační metody (diskriminační analýza lineární a kvadratická a jejich robustní varianty, metoda nejbližších sousedů a nejbližších prototypů, logistická regrese, probitová regrese, vícevrstvá perceptronová síť, klasifikační stromy a lesy). Simulovat neznámé parametry metod a jejich citlivost na přesnost modelu.

## 5.4. Materiál

Pro dosažení cíle práce byly využity údaje získávané každoročním dotazníkovým šetřením probíhajícím na Katedře účetnictví a financí, EF JU. Patří mezi ně mj. Rozvaha v plném rozsahu, Výkaz zisků a ztrát v plném rozsahu a Dotazník k zemědělským podnikům. Tyto údaje byly dostupné pro roky 2000 až 2010. Ve vzorku jsou obsaženy pouze zemědělské podniky, které vedou účetnictví (podvojně účetnictví) – až na dvě výjimky (fyzické osoby) se jedná o obchodní společnosti a družstva.

Na základě definice finanční tísně byly vybrány podniky, u kterých byly údaje dostupné v pěti po sobě následujících letech. Tato období budou dále v práci značena symbolem  $t$ ,  $t = \{0, 1, 2, 3, 4\}$ . Ohrožení finanční tísní je stanoveno podle výše uvedeného kritéria na základě hodnot ukazatelů v obdobích  $t = 2$ ,  $t = 3$  a  $t = 4$ . Na základě hodnot ukazatelů v obdobích  $t = 0$  a  $t = 1$  (a též dynamiky mezi těmito obdobími) bude ověřována schopnost predikce ohrožení finanční tísní. V tabulce (Tab. 10) jsou uvedeny četnosti podniků (a jejich rozdělení do skupin) pro tato pětiletá období.

Z tohoto přístupu vyplývá, že určitý podnik může být v materiálu obsažen jedenkrát či až sedmkrát (kdy období  $t = 0$  odpovídá rok 2000 až 2006). Ve vzorku je obsaženo 135 unikátních zemědělských podniků, přičemž každý podnik je zařazen v průměru 2,67krát (v různých obdobích), počet pozorování je tedy 361 (Příloha A). Přirozeně může být určitý podnik v některých obdobích zařazen ve skupině prosperujících podniků a v jiných obdobích jako podnik ohrožený finanční tísní.

Důvodem tohoto postupu výběru je fakt, že odvětví zemědělství je ovlivňováno řadou externích faktorů s vysokou meziroční variabilitou (počasí, oscilace cen produktů, zemědělská politika). Dá se předpokládat, že tento přístup zlepší generalizační schopnosti vytvořených modelů.

Tab. 10 Rozdělení souboru

Časový interval	Počet pozorování	Z toho ve finanční tísní
2000 - 2004	27	21 (77,8%)
2001 - 2005	31	23 (74,2%)
2002 - 2006	38	10 (26,3%)
2003 - 2007	65	8 (12,3%)
2004 - 2008	71	17 (23,9%)
2005 - 2009	72	22 (30,6%)
2006 - 2010	57	6 (10,5%)
Celkem	361	107 (29,6%)

Pramen: vlastní zpracování

### Vztah základního a výběrového souboru

Proporce mezi základním a výběrovým souborem bude uvedena pro rok 2009. V tomto roce bylo ve výběru 112 podniků, což je 3,37 % z evidovaných obchodních společností a družstev v zemědělství v České republice. V těchto podnicích bylo zaměstnáno 5,85 % všech pracovníků v zemědělství a tyto podniky obhospodařovaly 5,61 % zemědělské půdy (Tab. 11).

Tab. 11 Vztah základního a výběrového souboru

	Česká republika	Vzorek
Počet podniků (obch. spol. a družstva)	3 324	112
Počet pracovníků (v tisících)	120,2	7,026
Výměra zemědělské půdy (v tisících ha)	3 525	197,68

Pramen: Zpráva o stavu zemědělství (2010) a výběrový soubor

## 5.5. Softwarové vybavení

Pro aplikaci metod uvedených v teoretické části bude využit zejména statistický program Statistica ve verzi 8.0 od společnosti StatSoft; a dále tabulkový kalkulátor Microsoft Office Excel 2002 včetně Visual Basic for Applications. Odhady robustních diskriminačních funkcí budou provedeny ve statistickém programu R (verze 2.14.1), konkrétně s využitím knihovny rrcov pomocí funkcí Linda(), QdaCov() a predict(); metoda nejbližších prototypů pomocí knihovny pamr.

## 6. Řešení a výsledky

### 6.1. Využitelnost stávajících klasifikačních modelů

V této části bude ověřena využitelnost některých klasifikačních modelů uvedených v kapitole 4. Většina z těchto modelů byla vytvořena za jiným účelem než předpovídat ohrožení finanční tísní (podle výše uvedené definice), některé byly vytvořeny pro jiné ekonomiky a v dřívějších letech.

Do aplikace nejsou zařazeny modely, které např. vyžadují delší než dvouletou časovou řadu vstupů, obsahují absolutní ukazatele, příp. tyto ve stálých cenách určitého roku, využívající údajů kapitálového trhu, obsahující volné parametry či modely bez stanoveného pravidla pro zařazení do skupin. Výjimkou je model IN95, který byl použit bez posledního členu (podíl závazků po lhůtě splatnosti na výnosech). U tohoto modelu byly použity váhy určené pro odvětví zemědělství.

Většina modelů má v definici klasifikačního pravidla tzv. šedou zónu, tedy interval, ve kterém nelze pozorování jednoznačně zařadit. Právě z důvodu jednoznačnosti zařazení a sestavení klasifikační matice typu  $2 \times 2$  byly tyto intervaly rozděleny na dvě poloviny a pozorování z nich jsou přiřazeny bližší skupině. Zde je předpokladem, že intervaly šedé zóny byly vytvořeny symetricky kolem vypočteného dělicího bodu.

V tabulce (Tab. 12) jsou uvedeny výsledky vybraných modelů pro dvě období. Na základě klasifikačních matic je dopočtena falešná negativita ( $FNr$ ), falešná pozitivita ( $FPr$ ) a celková chyba klasifikace ( $ERr$ ). Ve všech modelech je celková chyba klasifikace v období  $t = 1$  nižší než v období  $t = 0$ . Model finančního zdraví pro operační program a Zmijewskiho model dosahují nejvyšší falešné negativity, téměř všechny podniky řadí do skupiny prosperujících podniků, a tedy celková chyba je blízká relativní četnosti podniků ohrožených finanční tísní. Nejnižší celkové chyby dosáhl model IN95, Tafflerův model a Gurčíkův model (vše pro  $t = 1$ ). IN95 a Tafflerův model mají velmi vysokou specifitu (při senzitivitě zhruba 45 %), naopak Gurčíkův model má velmi vysokou senzitivitu (94 % zároveň při dobré specifitě 65 %). Zde se projevují podmínky sestavení modelu, který je určen pro odhad budoucí prosperity slovenských zemědělských podniků.

Tab. 12 Spolehlivost existujících klasifikačních modelů

Model (autor)	Období	<i>ERr</i> (%)	<i>FNr</i> (%)	<i>FPr</i> (%)
IN95	$t = 0$	33,0	73,8	15,8
	$t = 1$	<b>19,7</b>	57,0	3,9
IN99	$t = 0$	68,7	0,9	97,2
	$t = 1$	68,1	0,0	96,9
IN01	$t = 0$	56,5	15,9	73,6
	$t = 1$	26,6	22,4	28,4
IN05	$t = 0$	56,5	16,8	73,2
	$t = 1$	26,3	23,4	27,6
Altman $Z'$	$t = 0$	53,7	35,5	61,4
	$t = 1$	44,3	28,0	51,2
Altman $Z''$	$t = 0$	31,6	85,1	9,1
	$t = 1$	27,4	78,5	5,9
Gurčík	$t = 0$	37,4	4,7	51,2
	$t = 1$	<b>26,0</b>	<b>5,6</b>	<b>34,7</b>
Zmijewski	$t = 0$	29,9	94,4	2,8
	$t = 1$	29,1	95,3	1,2
Taffler	$t = 0$	32,7	67,3	18,1
	$t = 1$	<b>19,4</b>	54,2	4,7
FZ OP	$t = 0$	30,2	96,3	2,4
	$t = 1$	29,1	95,3	1,2

Pramen: vlastní zpracování

## 6.2. Profilová analýza

Profilová analýza spočívá v nalezení optimální dělící hranice pro každý ukazatel, pro kterou je celková chyba klasifikace nejnižší. Procedura se liší pro ukazatele, s jejichž rostoucí hodnotou roste pravděpodobnost ohrožení finanční tísní (dále jen ukazatele nákladového charakteru) a pro ukazatele, s jejichž klesající hodnotou roste pravděpodobnost ohrožení finanční tísní (dále jen ukazatele výnosového charakteru).

Pro ukazatele výnosového typu je klasifikační pravidlo konstruováno

$$x \leq \theta \Rightarrow \hat{G} = 1,$$

$$x > \theta \Rightarrow \hat{G} = 0,$$

kde  $x$  je posuzovaná hodnota ukazatele,  $\theta$  je prahová hodnota,  $\hat{G}$  je odhad skupinové příslušnosti ( $\hat{G} = 1$  je skupina podniků ohrožených finanční tísni a  $\hat{G} = 0$  je skupina prosperujících podniků).

Pokud je  $\theta < \min x$ , pak jsou všechna pozorování zařazena do skupiny  $\hat{G} = 0$  a celková chyba se pak rovná relativní četnosti skupiny  $G = 1$ . Pokud je  $\theta \geq \max x$ , pak jsou všechna pozorování zařazena do skupiny  $\hat{G} = 1$  a celková chyba se pak rovná relativní četnosti skupiny  $G = 0$ .

Pro ukazatele nákladového typu platí, že

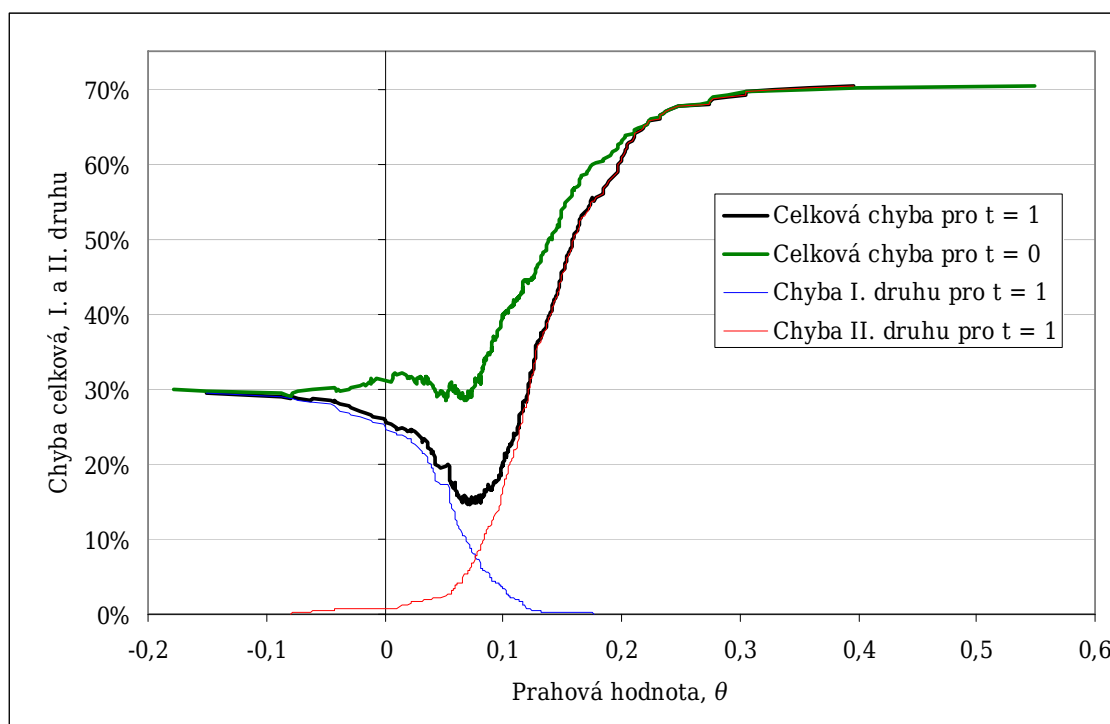
$$x \geq \theta \Rightarrow \hat{G} = 1,$$

$$x < \theta \Rightarrow \hat{G} = 0.$$

Pokud je  $\theta \leq \min x$ , pak jsou všechna pozorování zařazena do skupiny  $\hat{G} = 1$  a celková chyba se pak rovná relativní četnosti skupiny  $G = 0$ . Pokud je  $\theta > \max x$ , pak jsou všechna pozorování zařazena do skupiny  $\hat{G} = 0$  a celková chyba se pak rovná relativní četnosti skupiny  $G = 1$ .

Vývoj podílu chybné klasifikace v závislosti na prahové hodnotě u ukazatele  $CF/A$  je zobrazen v grafu (Obrázek 11). S rostoucí hodnotou  $\theta$  celková chyba nejprve klesá vlivem klesající chyby I. druhu až na hodnotu 14,68 %, dále roste vlivem rostoucí chyby II. druhu. Ukazatel  $CF/A$  je nejlepším klasifikátorem, pokud je prahová hodnota  $\theta = 0,0697$ . Celková chyba je pak rovna 14,68 %, tzn. že 308 podniků je zařazeno správně a 53 chybně, přičemž  $FNr = 31,8\%$  a  $FPr = 7,5\%$ .

Obrázek 11 Vývoj chyb u ukazatele cash rentabilita aktiv



Pramen: vlastní zpracování



Výsledky profilové analýzy zachycuje tabulka (Tab. 13). Je zde zařazeno 15 nejlepších ukazatelů podle celkové dosažené chyby klasifikace. Kromě celkové chyby je uvedena i hodnota falešné negativy a falešné positivity. Zde je patrné, jak celkovou chybu ovlivňují špatně klasifikovaná pozorování z obou skupin (před vážením rozsahy skupin). Nejnižší falešná negativita je dosažena u ukazatelů  $ZD/Mz_1$  a  $CF/A_1$ , kde je zhruba 31 % ohrožených podniků klasifikováno jako prosperující.

Tab. 13 Jednorozměrná klasifikační pravidla dle jejich spolehlivosti

Ukazatel	$ERr$ (%)	$FNr$ (%)	$FPr$ (%)	Prahová hodnota, $\theta$
$CF/A_1$ (+)	14,68	31,78	7,48	0,0697
$ZD/A_1$ (+)	14,96	33,64	7,09	0,0069
$ZD/PEP_1$ (+)	14,96	32,71	7,48	9,8776
$ZD/Mz_1$ (+)	15,79	30,84	9,45	0,0536
$ZUD/A_1$ (+)	16,07	39,25	6,30	0,0117
$ZD/V_1$ (+)	16,07	39,25	6,30	0,0070
$ZD/W_1$ (+)	16,07	38,32	6,69	0,0104
$CF/V_1$ (+)	16,62	38,32	7,48	0,0906
$DR_1$ (+)	18,01	42,06	7,87	0,0344
$CF/CZ_1$ (+)	19,39	36,45	12,20	0,1778
$ZD/W_{\delta'}$ (+)	19,67	53,27	5,51	-0,7412
$ZD/A_{\delta'}$ (+)	19,67	57,01	3,94	-0,9011
$ZD/V_{\delta'}$ (+)	19,67	57,01	3,94	-0,8913
$ZD/PEP_{\delta'}$ (+)	19,67	57,01	3,94	-0,8688
$ZD/Mz_{\delta'}$ (+)	19,67	57,01	3,94	-0,8994

Pramen: vlastní zpracování, (+) označuje ukazatele výnosového typu, (-) nákladového typu

### 6.3. Kombinace jednorozměrných klasifikačních pravidel

Z celkového počtu 108 klasifikačních pravidel je vybrána cca  $\frac{1}{4}$  (25) s nejnižší celkovou chybou klasifikace (Tab. 13). Vzhledem k faktu, že jsou některé ukazatele vzájemně velmi silně korelovány, je vhodné jejich počet snížit. Na základě matice spearmanových korelačních koeficientů (Tab. 14 zobrazuje již redukovanou matici) byl proveden užší výběr ukazatelů následujícím postupem.

1. Klasifikátory jsou seřazeny vzestupně podle dosaženého podílu chyb.
2. Do užšího výběru je zařazen nejlepší klasifikátor.

3. U dalšího klasifikátoru je posouzen jeho spearmanův korelační koeficient s každým již dříve zařazeným klasifikátorem. Pokud je  $-0,8 \leq r_s \leq 0,8$ , je klasifikátor zařazen do užšího výběru.

4. Krok 3 je zopakován pro každý další klasifikátor.

Tímto postupem bylo získáno 10 ukazatelů (Tab. 14), jejich spolehlivosti klasifikace jsou uvedeny v tabulce (Tab. 13). Obsaženy jsou ukazatele ze všech skupin kromě ukazatelů finanční stability a nákladovosti a třikrát též modifikovaná relativní změna ukazatele.

Tab. 14 Matice spearmanových korelačních koeficientů

		1	2	3	4	5	6	7	8	9	10
$CF/A_1$	1	1	0,76	0,43	0,66	0,45	0,40	0,44	0,41	-0,40	0,35
$ZD/PEP_1$	2	0,76	1	0,46	0,61	0,59	0,38	0,53	0,56	-0,33	0,15
$DR_1$	3	0,43	0,46	1	0,33	0,13	0,08	-0,03	0,36	-0,23	-0,04
$CF/CZ_1$	4	0,66	0,61	0,33	1	0,40	0,38	0,39	0,35	-0,20	0,16
$ZD/W_{\delta'}$	5	0,45	0,59	0,13	0,40	1	0,70	0,57	0,19	-0,16	0,11
$CF/CZ_{\delta'}$	6	0,40	0,38	0,08	0,38	0,70	1	0,50	0,14	-0,10	0,14
$DR_{\delta'}$	7	0,44	0,53	-0,03	0,39	0,57	0,50	1	0,20	-0,16	0,19
$V/PEP_1$	8	0,41	0,56	0,36	0,35	0,19	0,14	0,20	1	-0,33	0,22
$Zas/V_1$	9	-0,40	-0,33	-0,23	-0,20	-0,16	-0,10	-0,16	-0,33	1	-0,49
$V/A_1$	10	0,35	0,15	-0,04	0,16	0,11	0,14	0,19	0,22	-0,49	1

Pramen: vlastní zpracování

U 169 podniků bylo zařazení do skupiny správné u všech 10 ukazatelů. Na druhé straně 3 podniky (všechny ze skupiny ohrožené finanční tísni) nedokázal zařadit správně ani jeden z ukazatelů, 9 podniků (taktéž všechny ze skupiny ohrožené finanční tísni) zařadil správně pouze jeden ukazatel z 10 vybraných (více Tab. 15). Na tyto podniky nebude mít sestavení kombinace jednorozměrných klasifikátorů vliv a budou vždy klasifikovány nesprávně, chyba klasifikace tedy nikdy nepoklesne pod 3,3 %.

Tab. 15 Rozdělení podniků podle počtu správných klasifikací

Počet správných klasifikací	Počet podniků	Počet správných klasifikací	Počet podniků
0	3	6	19
1	9	7	22
2	16	8	25
3	11	9	58
4	13	10	169
5	14	Celkem	361

Pramen: vlastní zpracování

Dále bude ověřeno, zda skupina jednorozměrných klasifikátorů dosáhne nižší chyby klasifikace než nejlepší individuální klasifikátor. Z důvodu jednoznačného zařazení budou skupiny tvořeny vždy lichým počtem klasifikátorů. Z výše vybraných 10 ukazatelů tedy přicházejí v úvahu kombinace 3, 5, 7 a 9 ukazatelů. Počet těchto skupin je 120 pro kombinace 3 i 7 ukazatelů, 252 kombinací pěti ukazatelů a 10 kombinací devíti ukazatelů. Všechny ukazatele ve skupině budou mít stejnou váhu, ačkoliv by bylo možné uvažovat i o vahách rozdílných, např. podle individuální spolehlivosti.

### 6.3.1. Kombinace třech ukazatelů

Při vhodném výběru ukazatelů dojde ke zvýšení počtu správných klasifikací až na 314, tedy celková chyba klasifikace činí 13,02 %. Této chyby dosahují dvě kombinace ukazatelů, v obou jsou obsaženy zisk před zdaněním na pracovníka a cash rentabilita kapitálu (Tab. 16). Zlepšení klasifikační síly oproti nejlepšímu individuálnímu klasifikátoru ( $CF/A_1$ ) bylo způsobeno snížením falešné pozitivivity.

Tab. 16 Kombinace třech ukazatelů

$ERr$ (%)	Zahrnuté ukazatele	$FNr$ (%)	$FPr$ (%)
13,02	$ZD/PEP_1, DR_1, CF/A_1$	32,71	4,72
	$ZD/PEP_1, ZAS/V_1, CF/A_1$	33,64	4,33
13,30	$ZD/PEP_1, DR_1, V/PEP_1$	35,51	3,94
13,57	$ZD/PEP_1, DR_1, CF/CZ_1$	32,71	5,51
14,13	$ZD/PEP_1, CF/A_1, CF/CZ_1$	29,91	7,48
	$DR_1, CF/A_1, CF/CZ_1$	33,64	5,91

Pramen: vlastní zpracování

### 6.3.2. Kombinace pěti ukazatelů

Nejlepší kombinace pěti ukazatelů zvýší počet správně zařazených podniků na 313 (celková chyba klasifikace je tedy 13,3 %). Ve všech kombinacích pěti ukazatelů, které zvyšují přesnost klasifikace je obsažen zisk před zdaněním na pracovníka a ve většině i dlouhodobá rentabilita (Tab. 17). Stejně jako v předchozím případě je zvýšení spolehlivosti způsobeno snížením falešné pozitivivity při současném zhoršení falešné negativity.

Tab. 17 Kombinace pěti ukazatelů

<i>ERr</i> (%)	Zahrnuté ukazatele	<i>FNr</i> (%)	<i>FPr</i> (%)
13,30	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , Z <sub>as</sub> /V <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub>	34,58	4,33
13,85	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , V/PEP <sub>1</sub> , CF/A <sub>1</sub> , ZD/W <sub>δ'</sub>	36,45	4,33
14,13	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , V/A <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub>	37,38	4,33
	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , Z <sub>as</sub> /V <sub>1</sub> , CF/A <sub>1</sub> , ZD/W <sub>δ'</sub>	38,32	3,94
	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub> , ZD/W <sub>δ'</sub>	35,51	5,12
14,40	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , V/PEP <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub>	36,45	5,12
	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , V/PEP <sub>1</sub> , CF/CZ <sub>1</sub> , ZD/W <sub>δ'</sub>	38,32	4,33
	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub> , CF/CZ <sub>δ'</sub>	38,32	4,33
	ZD/PEP <sub>1</sub> , Z <sub>as</sub> /V <sub>1</sub> , V/PEP <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub>	36,45	5,12
	ZD/PEP <sub>1</sub> , Z <sub>as</sub> /V <sub>1</sub> , V/PEP <sub>1</sub> , CF/A <sub>1</sub> , ZD/W <sub>δ'</sub>	39,25	3,94
	ZD/PEP <sub>1</sub> , Z <sub>as</sub> /V <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub> , ZD/W <sub>δ'</sub>	37,38	4,72

Pramen: vlastní zpracování

### 6.3.3. Kombinace sedmi ukazatelů

Ze 120 kombinací sedmi ukazatelů pouze čtyři zvýšily počet správně klasifikovaných podniků. I v těchto kombinacích je obsažen zisk před zdaněním na pracovníka (Tab. 18). Podobně jako u kombinací tří a pěti ukazatelů i u kombinací sedmi ukazatelů je zlepšení spolehlivosti klasifikace způsobeno snížením falešné pozitivivity (při současném nárůstu falešné negativity).

Tab. 18 Kombinace sedmi ukazatelů

<i>ERr</i> (%)	Zahrnuté ukazatele	<i>FNr</i> (%)	<i>FPr</i> (%)
13,85	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , Z <sub>as</sub> /V <sub>1</sub> , V/PEP <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub> , ZD/W <sub>δ'</sub>	38,32	3,54
14,13	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , V/A <sub>1</sub> , V/PEP <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub> , ZD/W <sub>δ'</sub>	39,25	3,54
14,40	ZD/PEP <sub>1</sub> , DR <sub>1</sub> , V/A <sub>1</sub> , Z <sub>as</sub> /V <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub> , ZD/W <sub>δ'</sub>	41,12	3,15
	ZD/PEP <sub>1</sub> , V/A <sub>1</sub> , Z <sub>as</sub> /V <sub>1</sub> , V/PEP <sub>1</sub> , CF/A <sub>1</sub> , CF/CZ <sub>1</sub> , ZD/W <sub>δ'</sub>	41,12	3,15

Pramen: vlastní zpracování

### 6.3.4. Kombinace devíti ukazatelů

Kombinace devíti ukazatelů vznikly vyřazením vždy jednoho ukazatele. Při vyřazení modifikované relativní změny dlouhodobé rentability se celková chyba modelu snižuje na 12,7 % (Tab. 19).

Tab. 19 Kombinace devíti ukazatelů

$ERr$ (%)	Vyřazený ukazatel	$FNr$ (%)	$FPr$ (%)
12,74	$DR_{\delta'}$	33,64	3,94
13,57	$CF/CZ_{\delta'}$	34,58	4,72
13,57	$V/PEP_1$	35,51	4,33
13,85	$ZD/W_{\delta'}$	37,38	3,94
13,85	$DR_1$	36,45	4,33
13,85	$ZD/PEP_1$	38,32	3,54
14,40	$CF/A_1$	39,25	3,94
14,40	$Zas/V_1$	37,38	4,72
14,40	$V/A_1$	37,38	4,72
14,96	$CF/CZ_1$	40,19	4,33

Pramen: vlastní zpracování

#### Dílčí shrnutí

Zvýšení spolehlivosti klasifikace, které přinášejí klasifikační pravidla založená na kombinaci ukazatelů, je statisticky nevýznamné (14,68 % u modelu ukazatele  $CF/A_1$  oproti 12,74 % modelu devíti ukazatelů,  $p$ -level = 0,45) při současném nárůstu složitosti modelu. Navíc dochází ke snižování senzitivity modelů, tedy nedochází k včasnému varování o blížícím se ohrožení finanční tísní.

Při zařazení ukazatelů vztažených na jednoho pracovníka jsou prahové hodnoty závislé na cenové úrovni, jejich zařazení snižuje využitelnost výsledků v odlišných ekonomických podmínkách. Modifikované relativní změny ukazatelů přepočtených na pracovníka jsou ovlivněny pouze meziročním cenovým efektem.

## 6.4. Klasifikační síla ukazatelů

Klasifikační sílu ukazatelů je vhodné posuzovat pomocí ROC křivek (graficky) a z nich vycházejících veličin (numericky), tedy podle plochy pod ROC křivkou či podle minimální vzdálenosti od bodu [0, 1]. Klasifikační síla je pomocí těchto kvantitativních veličin u každého ukazatele posuzována v období  $t = 0$  a  $t = 1$  a

těž pro modifikovanou relativní změnu ukazatele mezi těmito obdobími (Tab. 20 uvádí 10 ukazatelů s nejvyšší hodnotou *AUC*).

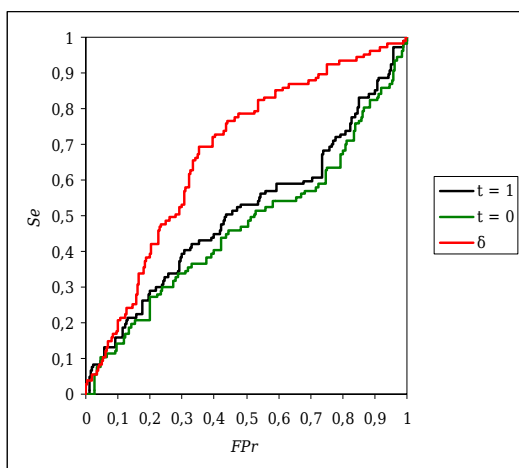
Tab. 20 Plochy pod ROC křivkou a minimální vzdálenosti od bodu [0, 1]

Ukazatel	Plocha pod ROC křivkou, <i>AUC</i>			Minimální vzdálenost od [0, 1]		
	$t = 1$	$t = 0$	$\delta'$	$t = 1$	$t = 0$	$\delta'$
ZD/A	0,8811	0,6579	0,7242	0,2863	0,5570	0,4762
ZUD/A	0,8826	0,6642	0,7147	0,2800	0,5377	0,4764
DR	0,8040	0,7452	0,7253	0,3692	0,4099	0,4374
ZD/V	0,8736	0,6492	0,7272	0,2880	0,5592	0,4853
ZD/W	0,8773	0,6506	0,7309	0,2880	0,5687	0,4764
Zas/V	0,7305	0,6623	0,6286	0,4594	0,5450	0,5754
ZD/Mz	0,8781	0,6531	0,7269	0,2924	0,5597	0,4846
CF/A	0,9069	0,7321	0,6804	0,2288	0,4279	0,4764
CF/V	0,8664	0,6910	0,6800	0,2793	0,4975	0,5023
CF/CZ	0,8083	0,6477	0,6912	0,3473	0,5178	0,4721

Pramen: vlastní zpracování

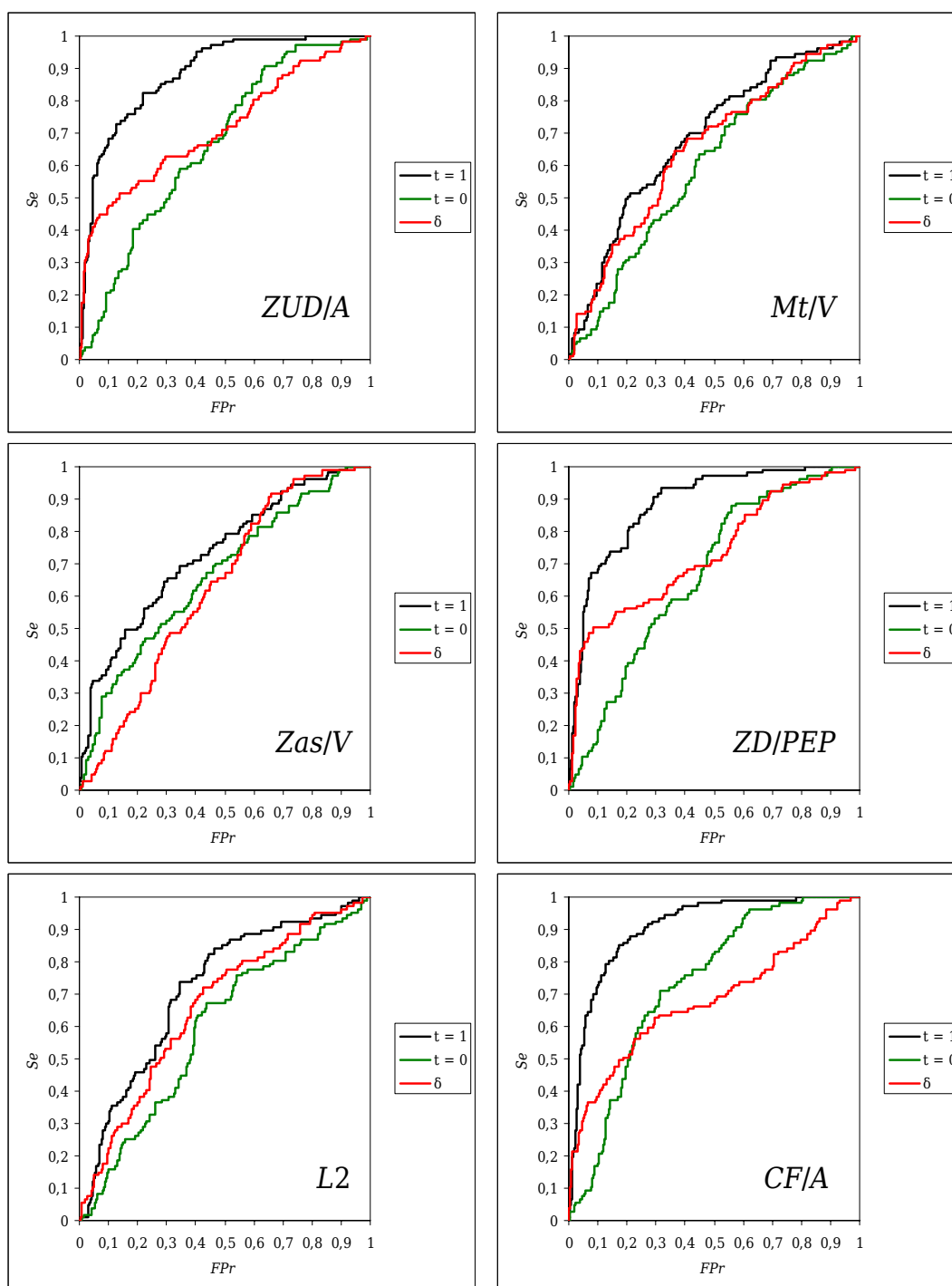
Obecně lze tvrdit, že klasifikátor je tím kvalitnější, čím je *AUC* vyšší nebo čím je vzdálenost od bodu [0, 1] nižší. Za nejlepší klasifikátory lze označit všechny ukazatele rentability a ukazatele založené na peněžních tocích, kde u ukazatele cash rentability aktiv je *AUC* = 0,91. Pro názornost budou uvedeny ROC křivky pro nejlepší ukazatel z každé skupiny (Obrázek 13). Za zmínku stojí ukazatel zadluženosti, který je v některé formě zastoupen ve většině existujících klasifikačních modelů. Zde má však v obou obdobích velmi nízkou klasifikační sílu (*AUC* v obou obdobích blízka 0,5). Lépe se jeví klasifikační síla modifikované relativní změny zadluženosti, u níž *AUC* = 0,68 (Obrázek 12).

Obrázek 12 ROC křivka pro CZ/A



Pramen: vlastní zpracování

Obrázek 13 ROC křivky vybraných ukazatelů



Pramen: vlastní zpracování

## 6.5. Výběr ukazatelů pro vícerozměrné klasifikační metody

V předchozí části byla vybrána skupina deseti ukazatelů, které dávaly nejlepší výsledky v profilové analýze a nebyly vzájemně velmi silně korelovány. Zde bu-

de vybrána skupina ukazatelů, které jsou nejlepšími klasifikátory a nebudou vzájemně velmi silně korelovány. Ukazatele z této skupiny budou využívány v dalších částech - ve vícerozměrných klasifikačních metodách, které neobsahují vlastní postupy výběru proměnných (či je případně obsahují, ale nelze je z určitých důvodů použít). Z výše uvedených důvodů nebudou do výběru zařazené ukazatele přepočtené na pracovníka. I v tomto případě je proveden předvýběr 25 ukazatelů (s nejvyšší hodnotou  $AUC$ ). Na základě matice spearmanových korelačních koeficientů je proveden užší výběr ukazatelů následujícím postupem.

1. Klasifikátory jsou seřazeny sestupně podle hodnoty  $AUC$ .
2. Do užšího výběru je zařazen nejlepší klasifikátor.
3. U dalšího klasifikátoru je posouzen jeho spearmanův korelační koeficient s každým již zařazeným klasifikátorem. Pokud je  $-0,8 \leq r_s \leq 0,8$ , je klasifikátor zařazen do užšího výběru.
4. Krok 3 je zopakován pro každý další klasifikátor.

Tímto postupem bylo vybráno 11 ukazatelů (Tab. 21), které jsou nejlepšími klasifikátory a zároveň nejsou vzájemně velmi silně korelovány. Nejsilnější pozitivní vazba je mezi  $CF/A_1$  a  $ZD/Mz_1$  ( $r_s = 0,77$ ), negativní vazba je nejsilnější mezi  $CF/A_1$  a  $Mt/V_1$  ( $r_s = -0,46$ ).

Tab. 21 Matice spearmanových korelačních koeficientů

		1	2	3	4	5	6	7	8	9	10	11
$CF/A_1$	1	1	0,77	0,66	0,43	0,52	0,45	-0,40	0,44	0,31	0,40	-0,46
$ZD/Mz_1$	2	0,77	1	0,60	0,45	0,36	0,61	-0,33	0,56	0,29	0,41	-0,35
$CF/CZ_1$	3	0,66	0,60	1	0,33	0,31	0,40	-0,20	0,39	0,47	0,38	-0,19
$DR_1$	4	0,43	0,45	0,33	1	0,36	0,13	-0,23	-0,03	0,38	0,08	-0,18
$CF/A_0$	5	0,52	0,36	0,31	0,36	1	-0,18	-0,30	0,02	0,31	-0,43	-0,26
$ZD/W_{6'}$	6	0,45	0,61	0,40	0,13	-0,18	1	-0,16	0,57	0,13	0,70	-0,22
$Zas/V_1$	7	-0,40	-0,33	-0,20	-0,23	-0,30	-0,16	1	-0,16	-0,03	-0,10	0,22
$DR_{6'}$	8	0,44	0,56	0,39	-0,03	0,02	0,57	-0,16	1	0,19	0,50	-0,23
$L2_1$	9	0,31	0,29	0,47	0,38	0,31	0,13	-0,03	0,19	1	0,05	-0,04
$CF/CZ_{6'}$	10	0,40	0,41	0,38	0,08	-0,43	0,70	-0,10	0,50	0,05	1	-0,15
$Mt/V_1$	11	-0,46	-0,35	-0,19	-0,18	-0,26	-0,22	0,22	-0,23	-0,04	-0,15	1

Pramen: vlastní zpracování

V příloze (Příloha B) jsou uvedeny základní statistické charakteristiky vybraných ukazatelů, a to jak v období  $t = 0$  a  $t = 1$ , tak pro jejich modifikovanou relativní změnu, vždy pro jednotlivé skupiny ( $G = 0$  a  $G = 1$ ). Příloha rovněž obsahuje jejich histogramy, ze kterých jsou patrné rozdíly v tvaru hustoty pravděpodobnosti mezi skupinami.



### 6.5.1. Grubbsův test

Pomocí Grubbsova testu odlehlých pozorování je testována nulová hypotéza, že vzorek neobsahuje žádná odlehlá pozorování. Test byl proveden u vybraných ukazatelů v obou obdobích a též pro modifikovanou relativní změnu zvláště pro skupinu podniků ohrožených finanční tísňí a pro skupinu podniků prosperujících. Výsledky (hodnoty testových kritérií a dosažené hladiny významnosti) jsou uvedeny v tabulce (Tab. 22). Relativní změny ukazatelů jsou nestandardizované veličiny a nulová hypotéza je ve všech případech zamítnuta s vysokou pravděpodobností. Na druhé straně vzorek není kontaminován odlehlými hodnotami u ukazatele materiálové nákladovosti, vázanosti zásob a dlouhodobé rentability (alespoň v některé skupině či období).

Tab. 22 Grubbsův test odlehlých pozorování

Ukazatel	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
<i>CF/A</i>	<b>4,30**</b>	<b>5,59***</b>	<b>3,88*</b>	<b>4,21**</b>	7,03***	7,27***
<i>ZD/Mz</i>	3,64*	4,96***	<b>3,34</b>	<b>4,66***</b>	5,34***	15,29***
<i>CF/CZ</i>	5,89***	6,54***	<b>4,57***</b>	<b>6,14***</b>	<b>5,53***</b>	<b>10,67***</b>
<i>DR</i>	3,08	3,99*	<b>2,99</b>	<b>3,31</b>	<b>8,15***</b>	<b>10,90***</b>
<i>ZD/W</i>	4,24**	6,29***	3,91*	9,74***	<b>5,42***</b>	<b>15,04***</b>
<i>Zas/V</i>	2,36	2,33	<b>2,96</b>	<b>10,06***</b>	7,72***	15,28***
<i>L2</i>	5,11***	4,44***	<b>7,59***</b>	<b>11,82***</b>	6,59***	8,82***
<i>Mt/V</i>	2,66	2,98	<b>2,90</b>	<b>2,99</b>	5,97***	8,76***

Pramen: vlastní zpracování; \* $p$ -level < 0,05; \*\* $p$ -level < 0,01; \*\*\* $p$ -level < 0,001

### 6.5.2. Test normality

Nulová hypotéza o shodě empirického a teoretického - normálního - rozdělení je ověřována pomocí Kolmogorovova-Smirnovova testu. Hodnoty testových kritérií a dosažené hladiny spolehlivosti jsou uvedeny v tabulce (Tab. 23). Hodnoty testového kritéria vyjadřují maximální hodnotu rozdílu mezi empirickou a teoretickou distribuční funkcí. U modifikovaných relativních změn je nulová hypotéza zamítnuta téměř u všech ukazatelů v obou skupinách. Vzhledem k výběru ukazatelů se zřetelem na jejich standardizaci nelze hypotézu o shodě tvaru rozdělení s normálním zamítnout u ukazatelů materiálová nákladovost, vázanost zásob, cash rentabilita aktiv a dlouhodobá rentabilita.

Tab. 23 Kolmogorovův-Smirnovův test normality

Ukazatel	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
<i>CF/A</i>	<b>0,104</b>	<b>0,081</b>	<b>0,134*</b>	<b>0,079</b>	0,230**	0,329**
<i>ZD/Mz</i>	0,181**	0,133**	<b>0,154*</b>	<b>0,110**</b>	0,248**	0,392**
<i>CF/CZ</i>	0,195**	0,133**	<b>0,130</b>	<b>0,170**</b>	<b>0,235**</b>	<b>0,349**</b>
<i>DR</i>	0,135*	0,057	<b>0,094</b>	<b>0,052</b>	<b>0,262**</b>	<b>0,366**</b>
<i>ZD/W</i>	0,195**	0,150**	0,160**	0,167**	<b>0,256**</b>	<b>0,396**</b>
<i>Zas/V</i>	0,053	0,088*	<b>0,066</b>	<b>0,045</b>	0,168**	0,326**
<i>L2</i>	0,262**	0,263**	<b>0,244**</b>	<b>0,173**</b>	0,239**	0,273**
<i>Mt/V</i>	0,069	0,038	<b>0,058</b>	<b>0,053</b>	0,112	0,136**

Pramen: vlastní zpracování; \* $p$ -level < 0,05; \*\* $p$ -level < 0,01

### 6.5.3. Testy shody středních hodnot

Předpokladem pro úspěšnou klasifikaci jednotlivých pozorování je co nejmenší překryv skupinových hustot pravděpodobnosti daného znaku. Provedme tedy jednorozměrné  $t$ -testy středních hodnot mezi skupinou podniků prosperujících ( $G = 0$ ) a ohrožených finanční tísni ( $G = 1$ ), a to v obou obdobích a také pro modifikovanou relativní změnu. Ve všech případech byla nejprve otestována hypotéza o shodě rozptylů a podle statistické významnosti (pro  $\alpha = 0,05$ ) byla následně vybrána příslušná varianta  $t$ -testu. Výsledky (hodnoty testových kritérií  $t$  a dosažené hladiny významnosti) jsou uvedeny v následující tabulce (Tab. 24). Střední hodnoty ukazatelů výnosového typu jsou bez výjimky nižší ve skupině ohrožených podniků, u ukazatelů nákladového typu platí opak. Rozdíly ve středních hodnotách jsou až na výjimky ( $L2_0$  a  $Zas/V_{\delta'}$ ) statisticky vysoce významné.

Tab. 24 Testy shody středních hodnot mezi  $G = 0$  a  $G = 1$ 

Ukazatel	$t = 0$	$t = 1$	$\delta'$
<i>CF/A</i>	<b>-7,564***</b>	<b>-14,523***</b>	-4,342***
<i>ZD/Mz</i>	-4,279***	<b>-11,951***</b>	-3,047**
<i>CF/CZ</i>	-3,368**	<b>-8,755***</b>	<b>-3,853***</b>
<i>DR</i>	-7,387***	<b>-10,164***</b>	<b>-4,260***</b>
<i>ZD/W</i>	-4,329***	-10,907***	<b>-3,027**</b>
<i>Zas/V</i>	4,430***	<b>7,813***</b>	0,710
<i>L2</i>	-1,260	<b>-5,916***</b>	-3,124**
<i>Mt/V</i>	2,748**	<b>5,627***</b>	3,404**

Pramen: vlastní zpracování; \* $p$ -level < 0,05; \*\* $p$ -level < 0,01; \*\*\* $p$ -level < 0,001

Pouze ukazatel  $CF/A$  byl do užšího výběru zařazen jak z období  $t = 0$ , tak  $t = 1$ . Přesto provedme jednorozměrné párové  $t$ -testy shody středních hodnot (tedy mezi středními hodnotami ukazatelů v období  $t = 0$  a  $t = 1$ ) pro skupinu prosperujících a pro skupinu podniků ohrožených finanční tísni i u ostatních ukazatelů (Tab. 25). Ve sloupci „Rozdíl“ je uveden rozdíl středních hodnot, tedy např.  $CF/A_1 - CF/A_0$ . ve sloupci  $t$  je uvedena hodnota testového kritéria a dosažená významnost. Z tabulky je patrné, že ve skupině prosperujících podniků se hodnoty ukazatelů výnosového typu v čase zvyšují (zlepšují) a hodnoty ukazatelů nákladového typu se v čase snižují (zlepšují); ve skupině ohrožených podniků je vývoj ukazatelů právě opačný. Rozdíly ve středních hodnotách ukazatelů jsou statisticky významné (pro  $\alpha = 0,05$ ) u všech ukazatelů v obou skupinách kromě vázanosti zásob u ohrožených podniků.

Tab. 25 Párové testy shody středních hodnot mezi  $t = 0$  a  $t = 1$

Ukazatel	G = 0		G = 1	
	Rozdíl	$t$	Rozdíl	$t$
$CF/A$	<b>0,0262</b>	<b>5,809***</b>	<b>-0,0130</b>	<b>-2,418*</b>
$ZD/Mz$	0,1524	6,395***	-0,0919	-2,814**
$CF/CZ$	0,0892	6,245***	-0,0464	-2,244*
$DR$	0,0384	10,170***	-0,0148	-2,893**
$ZD/W$	0,0520	6,237***	-0,0293	-2,514*
$Zas/V$	-0,0220	-3,158**	0,0092	1,834
$L2$	0,4525	2,474*	-0,2781	-2,541*
$Mt/V$	-0,0119	-3,998***	0,0112	2,236*

Pramen: vlastní zpracování; \* $p$ -level < 0,05; \*\* $p$ -level < 0,01; \*\*\* $p$ -level < 0,001

## 6.6. Diskriminační analýza

### 6.6.1. Lineární diskriminační analýza

Skutečné podíly podniků ohrožených finanční tísni a prosperujících podniků nejsou známy, budou tedy odhadnuty na základě relativních četností skupin ve výběrovém souboru. Relativní četnost podniků ohrožených finanční tísni je v tomto případě 29,6 %. Pro srovnání lze uvést výsledky z tiskové zprávy České kapitálové informační agentury (Čekia 2010), kde mezi rizikové podniky a podniky v úpadku v roce 2010 patřilo 32,6 % zemědělských podniků. Taktéž nejsou známy náklady mylné klasifikace, a tedy budou uvažovány shodné,  $z(1|2) = z(2|1)$ .

Nejprve bude provedena diskriminační analýza se zahrnutím všech 11 ukazatelů z užšího výběru. Hypotéza o vícerozměrné normalitě je na základě testování vícerozměrné šikmosti a špičatosti v obou skupinách zamítnuta s vysokou významností (ve skupině podniků ohrožených finanční tísni test na základě šikmosti  $1\,865 > \chi_{0,95}^2(286) = 326$ ; na základě špičatosti  $34,2 > u_{0,975} = 1,96$ ; ve skupině prosperujících podniků test na základě šikmosti  $16\,116 > \chi_{0,95}^2(286) = 326$ ; na základě špičatosti  $190 > u_{0,975} = 1,96$ ). Z řady simulačních studií vyplývá, že lineární diskriminační analýza je vhodnou metodou i při nesplnění požadavku na vícerozměrnou normalitu vstupních dat. Boxův  $M$  test hypotézy o shodě kovariančních matic je velmi citlivý právě na nenormalitu dat a vysoká hodnota testového kritéria pak signalizuje nerovnost kovariančních matic nebo nenormalitu dat nebo obojí, v tomto případě  $M\gamma = 9,95 > F_{0,95}(66, 144\,322) = 1,3$ .

Wilksova lambda tohoto modelu sestaveného na základě 11 proměnných je  $\Lambda = 0,5620$  a pomocí klasifikační funkce je správně zařazeno 87,3 % podniků (falešná negativita je 29 % a falešná pozitivita 5,9 %; ověřeno pomocí resubstituce). Některé z 11 proměnných nejsou z hlediska sestaveného modelu významné, je proto vhodné ověřit možnost zjednodušení modelu krokovou metodou výběru proměnných.

### **Kroková lineární diskriminační analýza**

Pomocí dopředné krokové LDA byly do modelu vybrány (z 11 ukazatelů) tyto ukazatele -  $CF/A_1$ ,  $DR_1$ ,  $Zas/V_1$  a  $L2_1$ . Algoritmus byl ukončen v pátém kroku, pomocí zpětné krokové LDA by byly vybrány stejné vstupní ukazatele jako u metody dopředné (stejně tak v případě pouhého přidávání či odebírání proměnných).

Testy hypotézy o shodě empirického rozdělení 4 ukazatelů s vícerozměrným normálním rozdělením tuto shodu zamítají s vysokou pravděpodobností (ve skupině podniků ohrožených finanční tísni test na základě šikmosti  $180 > \chi_{0,95}^2(20) = 31,4$ ; na základě špičatosti  $7,2 > u_{0,975} = 1,96$ ; ve skupině prosperujících podniků test na základě šikmosti  $553 > \chi_{0,95}^2(20) = 31,4$ ; na základě špičatosti  $28 > u_{0,975} = 1,96$ ). I v tomto případě je pomocí Boxova  $M$  testu zamítnuta hypotéza o shodě kovariančních matic,  $Mh = 80 > \chi_{0,95}^2(10) = 18,3$ .

Výsledná diskriminační funkce má tvar

$$\text{LDF} = 20,8903 CF/A_1 + 4,8497 DR_1 - 4,7564 Zas/V_1 + 0,1801 L2_1 - 0,4037$$

a klasifikační pravidlo pak

$$\text{pokud } \text{LDF} > 0, \text{ pak } \hat{G} = 0, \text{ jinak } \hat{G} = 1.$$

Ze srovnání normovaných koeficientů vyplývá, že nejvyšší vliv na hodnotu diskriminační funkce má ukazatel  $CF/A_1$ , dále  $DR_1$ ,  $Zas/V_1$  a  $L2_1$ ,  $\mathbf{b}^{*T} = (0,65; 0,45; -0,27; 0,18)$ .

Wilksova lambda se v tomto modelu zhoršila na  $\Lambda = 0,5753$ . Pomocí této funkce je správně zařazeno 87 % pozorování (Tab. 26), falešná negativita (29 %) je značně vyšší než falešná pozitivita (6,3 %).

Tab. 26 Klasifikační matice - LDA se 4 ukazateli

	$\hat{G} = 1$	$\hat{G} = 0$
$G = 1$	76	31
$G = 0$	16	238

Pramen: vlastní zpracování

Statistická významnost diskriminace testovaná pomocí Pressovy  $q$ -statistiky ukazuje na statisticky velmi vysokou spolehlivost diskriminace,  $q = 197,5 > \chi_{0,99}^2(1) = 6,63$ . Pro výběr o rozsahu  $n = 361$  a diskriminaci do dvou skupin je na hladině  $\alpha = 0,01$  statisticky významná diskriminační funkce, která správně zařadí již 205 pozorování (tedy pouhých 57 %). Z tohoto důvodu není významnost dalších výsledků pomocí Pressovy  $q$ -statistiky ověřována.

Pokud modifikujeme klasifikační pravidlo pro LDF a zvolíme za hraniční bod  $\theta = 0,583$  (v tomto bodu je celková chyba nejnižší), pak spolehlivost vzroste sice jen na 88,1 %, falešná pozitivita vzroste na 10,2 %, ale významně klesne falešná negativita (na 15,9 %,  $p = 0,02$ ). Oba hraniční body pro LDF jsou v grafu ROC křivek (Obrázek 14) zobrazeny.

Tímto postupem lze částečně obejít odhad nákladů mylné klasifikace. Předpokládejme, že korekce hraničního bodu odpovídá zohlednění rozdílných nákladů mylné klasifikace, tedy

$$\ln \frac{z(\hat{G} = 0 | G = 1)}{z(\hat{G} = 1 | G = 0)} = 0,583, \text{ a tedy}$$

$$\frac{z(\hat{G} = 0 | G = 1)}{z(\hat{G} = 1 | G = 0)} = 1,79.$$

V tomto případě by tedy náklady chyby I. druhu byly 1,8 krát vyšší než náklady chyby II. druhu. Pokud vyjdeme z výsledků Altman et al. (1977), kde autoři odhadují náklady chyby I. druhu na 0,7 a chyby II. druhu na 0,02, pak hraničním bodem je  $\theta = \ln(0,7 / 0,02) = 3,56$ . Zvýšení poměru nákladů mylné klasifikace by způsobilo zvýšení senzitivity modelu na 100 % (tedy všechny ohrožené podniky jsou zařazeny správně) a zároveň snížení specifity na pouhých 27 % (tedy 73 % prosperujících podniků by bylo označených za ohrožené). Celková chyba modelu by pak byla 51,2 %.

## 6.6.2. Kvadratická diskriminační analýza

Kvadratická diskriminační funkce odvozená na základě 11 vstupních ukazatelů dosahuje spolehlivosti klasifikace pouze 65,6 % - ačkoliv falešná negativita se snížila na pouhých 8,4 %, došlo k nárůstu (oproti lineární diskriminační funkci) falešné pozitivivity na 45,3 %. Přestože je snižování falešné negativivity žadáným efektem, je zde vykompenzováno enormním nárůstem falešné pozitivivity.

Je proto možné vyjít z výsledků krokového výběru proměnných u LDA a ověřit spolehlivost kvadratické diskriminační funkce odvozené na základě čtyř významných ukazatelů, tedy  $CF/A_1$ ,  $DR_1$ ,  $Zas/V_1$  a  $L2_1$ . Pro další případné využití jsou dále uvedeny inverzní matice k odhadnutým skupinovým kovariančním maticím (řádky a sloupce v uvedeném pořadí) a logaritmy jejich determinantů.

$$\Sigma_{G=1}^{-1} = \begin{bmatrix} 444,5267 & -21,4619 & 56,2367 & -4,4609 \\ -21,4619 & 33,3060 & -7,1850 & -1,3968 \\ 56,2367 & -7,1850 & 92,9913 & -2,2569 \\ -4,4609 & -1,3968 & -2,2569 & 0,9869 \end{bmatrix}$$

$$\Sigma_{G=0}^{-1} = \begin{bmatrix} 310,6518 & -15,6032 & 59,6386 & -0,1495 \\ -15,6032 & 38,4923 & 7,1177 & -0,5185 \\ 59,6386 & 7,1177 & 106,9762 & -0,3445 \\ -0,1495 & -0,5185 & -0,3445 & 0,2235 \end{bmatrix}$$

$$\ln|\Sigma_{G=1}| = -13,8166$$

$$\ln|\Sigma_{G=0}| = -12,3656$$

Podíl správně klasifikovaných pozorování je v tomto případě shodný s výsledkem lineární diskriminační analýzy, tedy 87 %. Došlo však ke zlepšení falešné negativivity na 18,7 % a ke zhoršení falešné pozitivivity na 10,6 % (Tab. 27). Rozhodnutí o výběru modelu by záviselo na ocenění nákladů mylné klasifikace a na toleranci vůči složitosti modelu (uchovávání kovariančních matic a operace s maticemi).

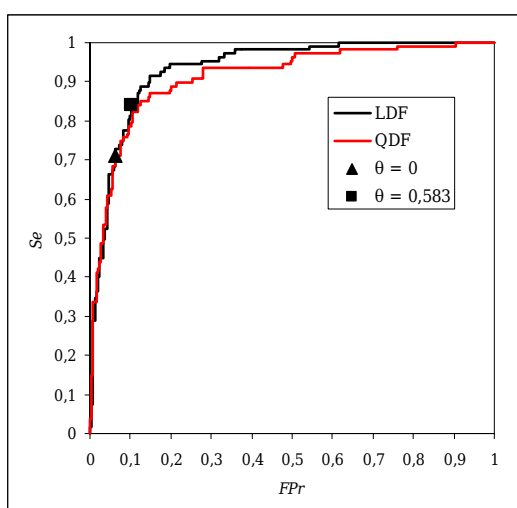
Tab. 27 Klasifikační matice - QDA se 4 ukazateli

	$\hat{G} = 1$	$\hat{G} = 0$
$G = 1$	87	20
$G = 0$	27	227

Pramen: vlastní zpracování

Vizuálně je možné posoudit kvalitu odhadnutých funkcí pomocí ROC křivek. V grafu (Obrázek 14) jsou vyneseny tyto křivky pro lineární a kvadratickou diskriminační funkci čtyř ukazatelů. Plocha pod ROC křivkou lineární diskriminační funkce je  $AUC = 0,9345$  a plocha pod ROC křivkou QDF je  $AUC = 0,9126$ .

Obrázek 14 ROC křivky diskriminačních funkcí



Pramen: vlastní zpracování

### 6.6.3. Robustní diskriminační analýza

#### Robustní lineární diskriminační analýza

Aplikace robustních odhadů polohy a měřítka v lineární diskriminační analýze bude provedena ve dvou variantách - pro 11 ukazatelů z užšího výběru a pro 4 ukazatele vybrané na základě krokové lineární diskriminační analýzy. U obou variant bude simulován různý předpokládaný podíl neodlehých pozorování, a to  $\alpha \in \{0,5; 0,6; 0,7; 0,8; 0,9\}$ .  $\alpha = 1$  odpovídá bodovým odhadům polohy a měřítka (tedy výsledkům „nerobustní - bodové“ diskriminační analýzy). Podíly chyb (Tab. 28 a Tab. 29) byly odhadnuty pomocí resubstituce. V případě 11 vstupních ukazatelů jsou všechny robustní modely z hlediska celkové chyby rovnocenné nebo horší než nerobustní model, modely pro  $\alpha = 0,7$  a  $\alpha = 0,8$  dosahují nejnižší falešné negativity.

Tab. 28 Podíl chyb - robustní LDA, 11 ukazatelů

$\alpha$	0,5	0,6	0,7	0,8	0,9	1
$ERr$ (%)	13,9	14,4	13,3	12,7	12,7	12,7
$FNr$ (%)	29,0	29,9	25,2	25,2	27,1	29,0
$FPr$ (%)	7,5	7,9	8,3	7,5	6,7	5,9

Pramen: vlastní zpracování

Při vstupních 4 ukazatelích je pro všechny  $\alpha$  falešná negativita téměř konstantní, celková chyba vychází nejnižší pro  $\alpha = 0,5$  a  $\alpha = 0,7$  vlivem snížené falešné pozitivivity. Je třeba si uvědomit, že v absolutním vyjádření se jedná o zlepšení pouze o 3 podniky.

Tab. 29 Podíl chyb - robustní LDA, 4 ukazatele

$\alpha$	0,5	0,6	0,7	0,8	0,9	1
<i>ERr</i> (%)	12,5	12,7	12,5	12,7	13,3	13,0
<i>FNr</i> (%)	29,9	29,9	29,9	29,9	29,0	29,0
<i>FPr</i> (%)	5,1	5,5	5,1	5,5	6,7	6,3

Pramen: vlastní zpracování

Pro ilustraci uvedme robustní lineární diskriminační funkci (např. pro  $\alpha = 0,7$ , kde je celková chyba 12,5 %). Výsledná robustní lineární diskriminační funkce má pak tvar

$$\text{RLDF} = 19,1124 \text{ CF/A}_1 + 5,1969 \text{ DR}_1 - 4,7343 \text{ Zas/V}_1 + \\ + 0,7738 \text{ L2}_1 - 0,9595$$

a klasifikační pravidlo pak

$$\text{pokud RLDF} > 0, \text{ pak } \hat{G} = 0, \text{ jinak } \hat{G} = 1.$$

Pokud porovnáme (i bez testování statistické významnosti rozdílů) lineární diskriminační funkci a robustní lineární diskriminační funkci z pohledu jejich parametrů a spolehlivosti, nutně dojdeme k závěru, že se liší jen minimálně. Tento závěr odpovídá faktu, že v diskriminačních funkcích jsou zařazeny převážně standardizované ukazatele (kromě likvidity), jejichž odhady středních hodnot a variability nejsou významně vychýleny vlivem odlehlých pozorování (viz Grubbsovy testy).

### Robustní kvadratická diskriminační analýza

V případě robustní kvadratické diskriminační analýzy je zvolen stejný postup jako v předchozím případě u robustní lineární diskriminační analýzy. V případě zapojení 11 ukazatelů (Tab. 30) se aplikací robustních odhadů celková chyba pro  $\alpha = 0,7$  snížila na 21,9 %, přesto zůstává nepříjemně vysoká. Při zapojení pouze ukazatelů  $\text{CF/A}_1$ ,  $\text{DR}_1$ ,  $\text{Zas/V}_1$  a  $\text{L2}_1$  (Tab. 31) nepřináší použití robustních odhadů požadovaný výsledek ve snížení celkové chyby (falešná pozitivita se zlepšuje na úkor falešné negativity).

Tab. 30 Podíl chyb - robustní QDA, 11 ukazatelů

$\alpha$	0,5	0,6	0,7	0,8	0,9	1
<i>ERr</i> (%)	22,2	24,9	21,9	27,4	22,7	34,3
<i>FNr</i> (%)	19,6	18,7	18,7	17,8	19,6	8,4
<i>FPr</i> (%)	23,2	27,6	23,2	31,5	24,0	45,3

Pramen: vlastní zpracování



Tab. 31 Podíl chyb - robustní QDA, 4 ukazatele

$\alpha$	0,5	0,6	0,7	0,8	0,9	1
<i>ERr</i> (%)	14,7	14,4	14,7	14,7	13,6	13,0
<i>FNr</i> (%)	29,0	28,0	29,0	29,0	28,0	18,7
<i>FPr</i> (%)	8,7	8,7	8,7	8,7	7,5	10,6

Pramen: vlastní zpracování

### Dílčí shrnutí

Pomocí krokové lineární diskriminační analýzy byly vybrány ukazatele  $CF/A_1$ ,  $DR_1$ ,  $Zas/V_1$  a  $L2_1$ . Ze srovnání normovaných koeficientů diskriminační funkce vyplývá, že nejvyšší vliv na hodnotu diskriminační funkce má ukazatel  $CF/A_1$ , dále  $DR_1$ ,  $Zas/V_1$  a  $L2_1$ . Pomocí lineární diskriminační funkce čtyř ukazatelů je správně zařazeno 87 % pozorování, přičemž falešná negativita (29 %) podstatně převyšuje falešnou pozitivitu (6,3 %). Pokud modifikujeme klasifikační pravidlo pro LDF a zvolíme za hraniční bod  $\theta = 0,583$  namísto  $\theta = 0$ , potom falešná pozitivita vzroste na 10,2 % při současném významném poklesu falešné negativy (15,9 %). Kvadratická diskriminační funkce sestavená na základě čtyř uvedených ukazatelů dosahuje shodné spolehlivosti s lineární diskriminační funkcí. Došlo však ke zlepšení falešné negativy na 18,7 % a ke zhoršení falešné pozitivy na 10,6 %.

Zapojení robustních odhadů parametrů nepřináší požadované výsledky ve snížení celkové chyby. Při výběru ukazatelů byl kladen důraz na jejich standardizaci, a tudíž jejichž odhady středních hodnot a variability nejsou významně vychýleny vlivem odlehlých pozorování.

Jak v lineární, tak v kvadratické diskriminační funkci nebyly brány v úvahu náklady mylné klasifikace. Je nutné zohlednit i složitost modelu, kdy u QDF je pro další využití nutno uchovávat vyšší počet parametrů.

## 6.7. Logistická regrese

První varianta modelu odvozeného na základě logistické regresní analýzy předpokládá zařazení všech 11 ukazatelů z užšího výběru. Modelovanou pravděpodobností je budoucí prosperita podniku, tedy  $P(G = 0)$ .

Testy významnosti jednotlivých nezávislých proměnných založené na Waldově kritériu prokazují na hladině  $\alpha = 0,05$  významnost pouze 3 ukazatelů -  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$ , dále by bylo možné uvažovat o ukazateli  $L2_1$ , kde  $p$ -level = 0,08 (Tab. 32).

Tab. 32 Logistická regrese - odhady parametrů, 11 ukazatelů

	Odhad parametru	Směrodatná chyba	Waldovo kritérium	p-level
Abs. člen	-0,26319	1,337	0,039	0,844
$CF/A_1$	24,47931	7,588	10,408	0,001
$ZD/Mz_1$	1,57863	1,298	1,479	0,224
$CF/CZ_1$	-0,07373	0,967	0,006	0,939
$DR_1$	4,82468	1,148	17,673	0,000
$CF/A_0$	-1,26820	4,414	0,082	0,774
$ZD/W_{\delta'}$	-0,00689	0,017	0,350	0,554
$Zas/V_1$	-5,81042	1,751	11,017	0,001
$DR_{\delta'}$	0,02645	0,105	0,064	0,801
$L2_1$	0,26933	0,154	3,066	0,080
$CF/CZ_{\delta'}$	0,12549	0,142	0,777	0,378
$Mt/V_1$	-1,07189	2,721	0,155	0,694

Pramen: vlastní zpracování

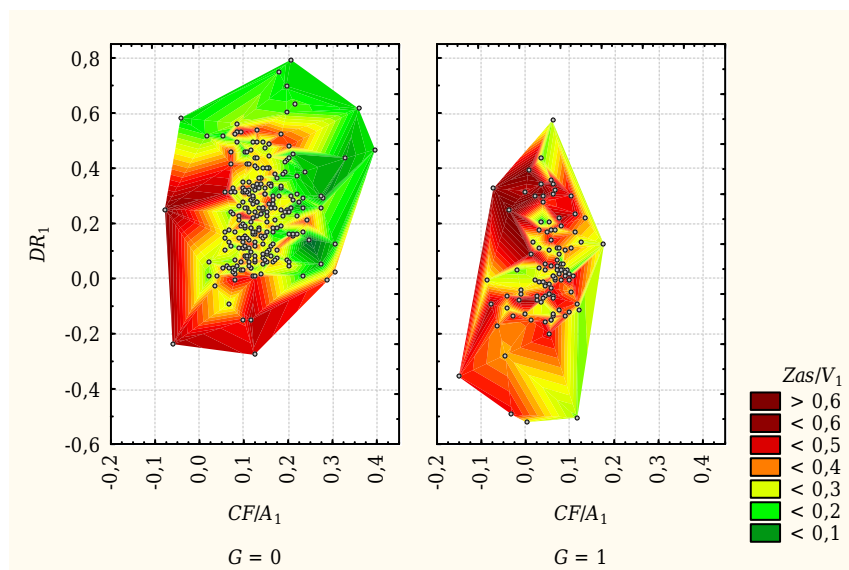
Věrohodnostní skóre tohoto modelu je  $LR = 158,11$ . Pokud je jako hranice pro klasifikaci uvažována hodnota  $\theta = 0,5$ , je pomocí něj správně zařazeno (ověřeno pomocí resubstituce) 87,5 % pozorování, falešná negativita činí 26,2 % a falešná pozitivita 6,7 %.

Kromě posouzení statistické významnosti ukazatelů je vhodné posoudit hodnoty jejich regresních koeficientů (s ohledem na specifika logistické regrese). V tomto případě pozitivní hodnoty koeficientu značí přímou závislost mezi hodnotou daného ukazatele a šancí, že je podnik prosperující. Znaménka parametrů u některých ukazatelů ( $CF/A_0$ ,  $CF/CZ_1$ ,  $ZD/W_{\delta'}$ ) lze označit přinejmenším za neintuitivní.

Dále tedy bude sestaven model pouze na základě 3 statisticky významných ukazatelů, tedy  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$ . Jejich hodnoty pro všechna pozorování je možné přehledně zobrazit pomocí wafer grafu (Obrázek 15). Z něj jsou patrné rozdíly mezi skupinami podniků v jejich prostorové distribuci. Nízké hodnoty vázanosti zásob dosahují pouze prosperující podniky, a to zejména s vysokou dlouhodobou a vysokou cash rentabilitou aktiv. Ukazatel vázanosti zásob se z modelů uvedených v části Klasifikační modely vyskytuje pouze v modelu pro odhad budoucí prosperity zemědělských podniků (G-index; Gurčík 2002) a v Indexu bonity (Kralicek 1993). Kromě těchto prací se ukazatel vázanosti zásob (konkrétně ve formě doby obratu zásob) projevil jako významný v práci Jakubík a Teplý (2007) při předpovědi úpadku českých podniků. Tento výsledek je pro autory studie překvapivý i vzhledem k faktu, že jejich vstupní soubor obsahoval podniky všech odvětví národního hospodářství, přitom právě hodnoty tohoto ukazatele se mezi odvětvími značně liší. Ukazatel dlouhodobé rentability je obsažen ve všech Altmanových modelech, v G-indexu a v modelu finančního

zdraví pro Operační program a Program rozvoje venkova (Rosochatecká a Řezbová 2004).

Obrázek 15 Wafer graf  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$



Pramen: vlastní zpracování, STATISTICA 8

Tab. 33 Logistická regrese - odhady parametrů, 3 ukazatele

	Odhad parametru	Směrodatná chyba	Waldovo kritérium	<i>p</i> -level
Abs. člen	-1,01032	0,733	1,899	0,168
$CF/A_1$	32,45628	4,686	47,968	0,000
$DR_1$	5,26484	1,052	25,020	0,000
$Zas/V_1$	-5,24527	1,635	10,284	0,001

Pramen: vlastní zpracování

Věrohodnostní skóre tohoto modelu je  $LR = 150,45$ . Pokud je jako hranice pro klasifikaci uvažována hodnota  $\theta = 0,5$ , je pomocí něj správně zařazeno (ověřeno pomocí resubstituce) 87 % pozorování (Tab. 34), oproti plnému modelu klesla falešná negativita ( $FNr = 25,2\%$ ) a vzrostla falešná pozitivita ( $FPr = 7,9\%$ ).

Tab. 34 Klasifikační matice - logistická regrese, 3 ukazatele

	$\hat{G} = 1$	$\hat{G} = 0$
$G = 1$	80	27
$G = 0$	20	234

Pramen: vlastní zpracování

Kvalitu modelu je vhodné posoudit sestavením ROC křivky a pomocí číselných charakteristik této křivky. Tyto křivky jsou vyneseny do grafu (Obrázek 16) a jejich tvar dokládá výtečné klasifikační schopnosti logistických modelů. Plocha

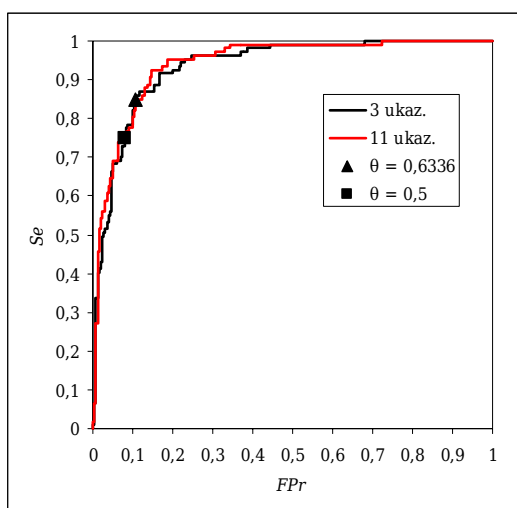
pod křivkou modelu s 11 ukazateli je  $AUC = 0,9415$  a modelu s 3 ukazateli je  $AUC = 0,9345$ .

Výše uvedené spolehlivosti klasifikace a klasifikační matice byly vypočteny na základě klasifikačního pravidla

$$P(G = 0) > \theta = 0,5 \Rightarrow \hat{G} = 0.$$

Pokud se liší rozsahy souborů nebo náklady mylné klasifikace, je vhodné hledat jinou prahovou hodnotu namísto  $\theta = 0,5$ . V případě modelu se 3 ukazateli je touto hodnotou  $\theta = 0,6336$ , pro kterou je celková chyba nejnižší. Celková spolehlivost vzroste sice jen na 88,1 %, k výraznější změnám však dojde u falešné pozitivivity a falešné negativity. Zatímco falešná pozitivita vzroste na 10,6 %, falešná negativita poklesne na 14,95 % (statistická významnost tohoto snížení je  $p = 0,062$ ).

Obrázek 16 ROC křivky logistických modelů



Pramen: vlastní zpracování

### Dílčí shrnutí

Při tvorbě modelu pomocí logistické regrese byla modelovanou pravděpodobností budoucí prosperita podniku, tedy  $P(G = 0)$ . Testy významnosti Waldova kritéria jednotlivých nezávislých proměnných prokazují významnost pouze ukazatelů  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$  (na hladině  $\alpha = 0,05$ ).

Pokud je jako hranice pro klasifikaci uvažována hodnota  $\theta = 0,5$ , je pomocí modelu tří ukazatelů správně zařazeno 87 % pozorování, falešná negativita je 25,2 % a falešná pozitivita je 6,7 %. Pokud se liší rozsahy souborů nebo náklady mylné klasifikace, je vhodné hledat jinou prahovou hodnotu. Model se 3 ukazateli dosahuje nejnižší celkové chyby pro  $\theta = 0,6336$ . Celková spolehlivost při takto upraveném klasifikačním pravidle vzroste sice jen na 88,1 %, k výraznější změnám však dojde u falešné pozitivivity a falešné negativity. Zatímco falešná pozitivita vzroste na 10,6 %, falešná negativita poklesne na 14,95 %.

## 6.8. Probitová regrese

I u metody probitové regrese uvažujeme nejprve o modelu, který zahrnuje všech 11 ukazatelů užšího výběru. Modelovanou pravděpodobností je stejně jako u logistického modelu budoucí prosperita podniku, tedy  $P(G = 0)$ .

Stejně jako u předchozí logistické regrese testy významnosti jednotlivých nezávislých proměnných založené na Waldově kritériu prokazují na hladině  $\alpha = 0,05$  významnost pouze 3 ukazatelů -  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$  (Tab. 35).

Tab. 35 Probitová regrese - odhady parametrů, 11 ukazatelů

	Odhad parametru	Směrodatná chyba	Waldovo kritérium	p-level
Abs. člen	0,17646	0,747	0,055	0,813
$CF/A_1$	10,97346	3,802	8,327	0,003
$ZD/Mz_1$	0,87913	0,666	1,738	0,187
$CF/CZ_1$	-0,01102	0,510	0,0005	0,982
$DR_1$	2,71784	0,614	19,546	0,000
$CF/A_0$	-0,77429	2,363	0,107	0,743
$ZD/W_{\delta'}$	-0,00191	0,007	0,076	0,781
$Zas/V_1$	-2,80766	0,937	8,963	0,002
$DR_{\delta'}$	0,02272	0,054	0,175	0,675
$L2_1$	0,13077	0,081	2,550	0,110
$CF/CZ_{\delta'}$	0,06534	0,073	0,780	0,376
$Mt/V_1$	-1,04022	1,483	0,491	0,483

Pramen: vlastní zpracování

Věřohodnostní skóre tohoto probitového modelu založeného na 11 ukazatelích je  $LR = 158,11$  (stejně jako u logistického modelu). Pokud je jako hranice pro klasifikaci uvažována hodnota  $\theta = 0,5$ , je pomocí něj správně zařazeno (ověřeno pomocí resubstituce) 87 % pozorování, falešná negativita činí 29 % a falešná pozitivita 6,3 %. Znaménka některých regresních koeficientů (u ukazatelů  $CF/A_0$ ,  $CF/CZ_1$ ,  $ZD/W_{\delta'}$ ) lze i v tomto případě označit za neintuitivní.

Dále tedy bude sestaven probitový model na základě pouze 3 statisticky významných ukazatelů, tedy  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$  (Tab. 36).

Tab. 36 Probitová regrese - odhady parametrů, 3 ukazatele

	Odhad parametru	Směrodatná chyba	Waldovo kritérium	p-level
Abs. člen	-0,45471	0,404	1,262	0,261
$CF/A_1$	15,78512	2,273	48,204	0,000
$DR_1$	2,88674	0,562	26,343	0,000
$Zas/V_1$	-2,53790	0,889	8,136	0,004

Pramen: vlastní zpracování

Věrohodnostní skóre redukovaného probitového modelu je  $LR = 150,45$ . Pokud je jako hranice pro klasifikaci uvažována hodnota  $\theta = 0,5$ , je pomocí něj správně zařazeno (ověřeno pomocí resubstituce) 86,7 % pozorování (Tab. 37), oproti plnému byl navíc mylně klasifikován 1 ohrožený podnik jako prosperující, a tedy vzrostla falešná negativita ( $FNr = 29,9 \%$ ), specifita (a tedy i falešná pozitivita) zůstala bez změny.

Tab. 37 Klasifikační matice - probitová regrese, 3 ukazatele

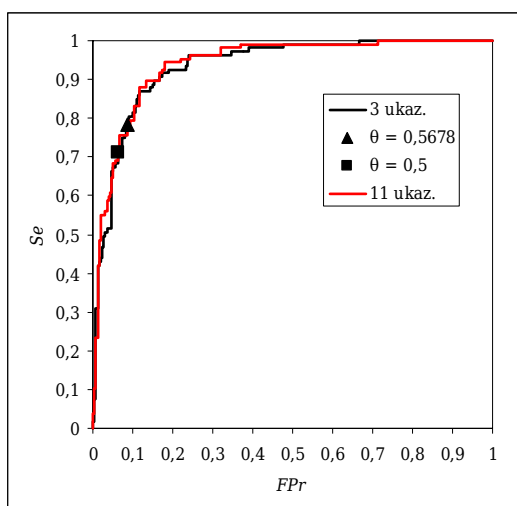
	$\hat{G} = 1$	$\hat{G} = 0$
$G = 1$	75	32
$G = 0$	16	238

Pramen: vlastní zpracování

ROC křivky probitových modelů jsou zobrazeny v grafu (Obrázek 17). Tvar ROC křivek probitových modelů dokládá jejich výtečné klasifikační schopnosti. Plocha pod křivkou modelu s 11 ukazateli je  $AUC = 0,94$  a modelu s 3 ukazateli je  $AUC = 0,934$ .

Hledáním jiného prahového bodu (než  $\theta = 0,5$ ), který by zvyšoval pravděpodobnost správné klasifikace, dojdeme k výsledku  $\theta = 0,5678$  (v případě modelu se 3 ukazateli). Celková spolehlivost vzroste sice jen na 87,8 %, ale falešná negativita poklesne na 20,6 % a falešná pozitivita vzroste na 8,7 %. Oba prahové body jsou zaneseny v grafu.

Obrázek 17 ROC křivky probitových modelů



Pramen: vlastní zpracování

### Dílčí shrnutí

Probitová regrese je velmi podobná logistické regresi jak metodicky, tak i výsledky. Obdobně jako u logistické regrese testy významnosti jednotlivých nezávislých proměnných založené na Waldově kritériu prokazují na hladině  $\alpha = 0,05$  významnost pouze ukazatelů  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$ . Pokud je hranice pro klasifikaci stanovena  $\theta = 0,5$ , je pomocí probitového modelu správně zařazeno 86,7 % pozorování, falešná negativita je 29,9 % a falešná pozitivita 6,3 %. Při hranici  $\theta = 0,5678$  vzrostla celková spolehlivost na 87,8 %, falešná negativita poklesla na 20,6 % a falešná pozitivita vzrostla na 8,7 %.

## 6.9. Metoda nejbližších sousedů

### 6.9.1. Metoda nejbližšího souseda

Výsledky metody nejbližších sousedů byly získány na základě vlastní algoritmi-zace ve VBA MS Excel. Schopnosti klasifikace byly ověřeny u 13 modelů, a to pomocí metody kapesního nože (jack-knife cross validation) - byla tedy porov-návána vzdálenost (euklidovská) každého pozorování se všemi ostatními. Před výpočtem vzdáleností byla vstupní data normalizována metodou směrodatné proměnné.

Tabulka (Tab. 38) obsahuje výsledky (dosažená celková chyba, falešná negativita a falešná pozitivita) třinácti modelů. Prvních 11 z nich bylo konstruováno na základě Tab. 21 následovně. První model obsahoval pouze ukazatel  $CF/A_1$ , dru-hý obsahoval  $CF/A_1$  a  $ZD/Mz_1$ , ve třetím byl k těmto dvěma ukazatelům přidán

ukazatel  $CF/CZ_1$  atd. Poslední dva modely obsahují kombinace ukazatelů, které byly pomocí krokové diskriminační analýzy a krokové logistické regrese označeny za významné.

Tab. 38 Metoda nejbližšího souseda - dosažená spolehlivost

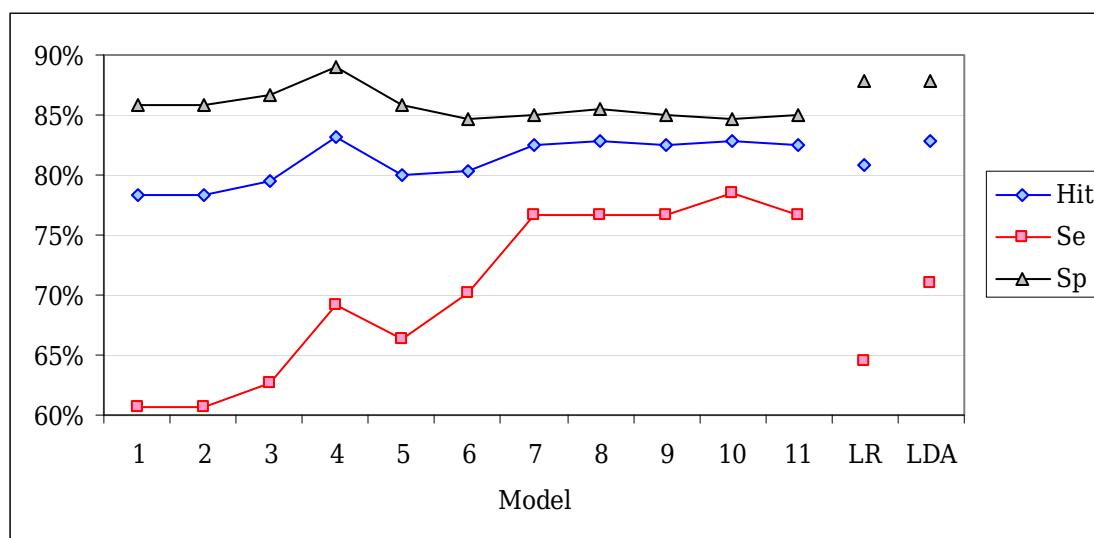
Ukazatel	Model	$ERr$ (%)	$FNr$ (%)	$FPr$ (%)
$CF/A_1$	1	21,6	39,3	14,2
$ZD/Mz_1$ a výše	2	21,6	39,3	14,2
$CF/CZ_1$ a výše	3	20,5	37,4	13,4
$DR_1$ a výše	4	<b>16,9</b>	<b>30,8</b>	<b>11,0</b>
$CF/A_0$ a výše	5	19,9	33,6	14,2
$ZD/W_{\delta}$ a výše	6	19,7	29,9	15,4
$Zas/V_1$ a výše	7	17,5	23,4	15,0
$DR_{\delta}$ a výše	8	17,2	23,4	14,6
$L2_1$ a výše	9	17,5	23,4	15,0
$CF/CZ_{\delta}$ a výše	10	17,2	21,5	15,4
$Mt/V_1$ a výše	11	17,5	23,4	15,0
$CF/A_1, DR_1, Zas/V_1$	LR	19,1	35,5	12,2
$CF/A_1, DR_1, Zas/V_1, L2_1$	LDA	<b>17,2</b>	<b>29,0</b>	<b>12,2</b>

Pramen: vlastní zpracování

Výsledky z tabulky (Tab. 38) jsou názorně zopakovány (ve formě senzitivity, specifity a celkové spolehlivosti) v grafu (Obrázek 18). Se složitostí modelu (s počtem vstupních ukazatelů) dochází k zlepšování senzitivity (tedy správné klasifikace ohrožených podniků), a to až na úroveň 78,5 %. Celková spolehlivost je nejvyšší u modelu čtyř ukazatelů ( $CF/A_1, ZD/Mz_1, CF/CZ_1$  a  $DR_1$ ), a to 83,1 %, a dobré výsledky (spolehlivost vs. složitost) též podává model ukazatelů odvozených v krokové lineární diskriminační analýze.



Obrázek 18 Metoda nejbližšího souseda - dosažená spolehlivost



Pramen: vlastní zpracování

Jak bylo uvedeno výše, jako míra podobnosti objektů byla užita euklidovská vzdálenost. Při použití jiné metody obdržíme odlišné výsledky. Pro ilustraci např. při použití Hemmingovy vzdálenosti (Manhattan city blocks) pro model 4 ukazatelů vybraných při krokové LDA se výsledky mírně liší ( $ERr = 16,9\%$ ,  $FNr = 29,9\%$  a  $FPr = 11,4\%$ ).

## 6.9.2. Metoda $k$ -nejbližších sousedů

Při použití metody  $k$ -nejbližších sousedů (pro  $k > 1$ ) je nutné zvolit počet nejbližších sousedů, tedy hodnotu  $k$ . Pro tento účel bude posouzen vývoj celkové chyby, falešné positivity a falešné negativity pro  $k \in \{1, 2, \dots, 40\}$ . Vzhledem k faktu, že rozsahy skupin se liší, je nutno při konstrukci klasifikačního pravidla relativní četnosti skupin zohlednit.

Pozorování je zařazeno do skupiny, pro kterou je podíl  $k_i / n_i$  nejvyšší; kde  $k_i$  je počet nejbližších sousedů ze skupiny  $i$  a  $n_i$  je četnost  $i$ -té skupiny.

Ilustrační příklad:  $k = 20$ ,  $k_1 = 8$ ,  $k_0 = 12$ ,  $n_1 = 107$ ,  $n_0 = 254$

$$k_1 / n_1 = 8 / 107 = 0,0748$$

$$k_0 / n_0 = 12 / 254 = 0,0472$$

Řešené pozorování bude zařazeno do skupiny 1. Hyperkoule sice obsahuje více pozorování ze skupiny 0 (60%), ovšem vztaheno k rozsahům skupin je poměr vyšší pro skupinu 1.

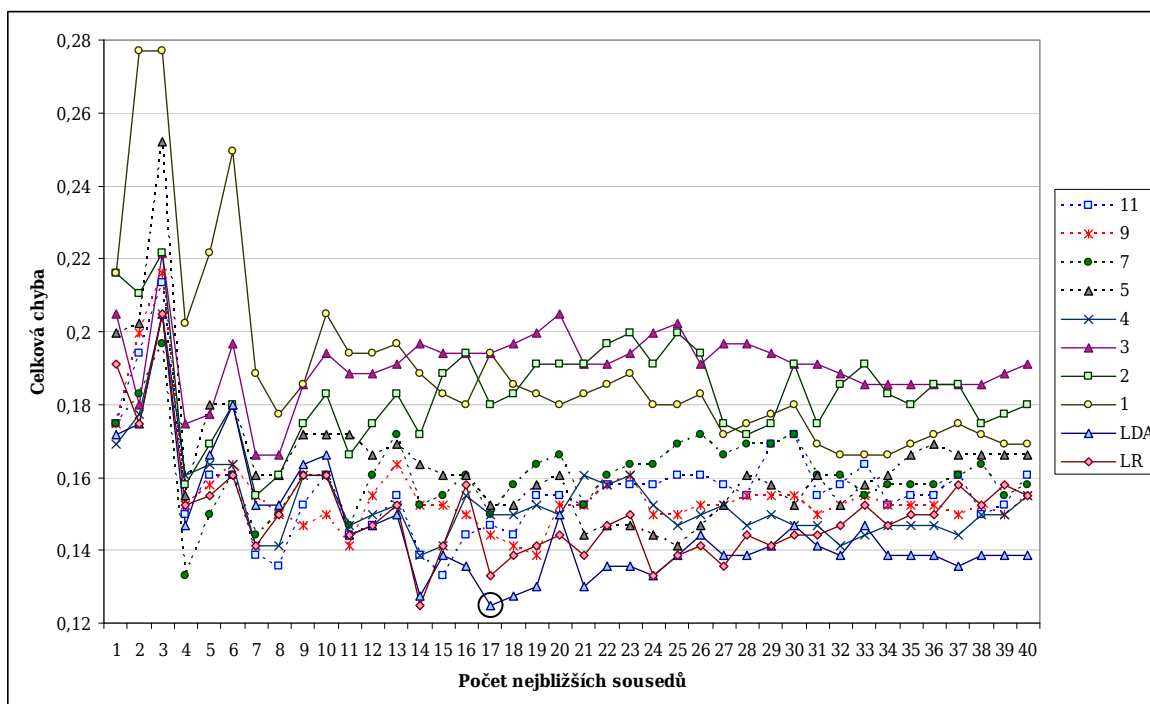
Stejně jako v předchozí úloze (1-NN) bude uvažováno 13 modelů. Jedenáct z nich bude konstruováno na základě Tab. 21 (stejným postupem jako v případě 1-NN) a budou doplněny o dva modely obsahující kombinace ukazatelů, které byly pomocí krokové diskriminační analýzy a krokové logistické označeny za významné. Kvalita klasifikace byla ověřována pomocí metody kapesního nože,

klasifikováno bylo každé pozorování. Další možností by bylo ověřování na náhodně vybraném testovacím souboru či křížové ověřování.

V následujících grafech nejsou z důvodu nízké přehlednosti uvedeny všechny modely. Obrázek 19 zobrazuje vývoj celkové chyby modelu pro rostoucí počet nejbližších sousedů, Obrázek 20 znázorňuje vývoj falešné negativy a Obrázek 21 vývoj falešné pozitivy.

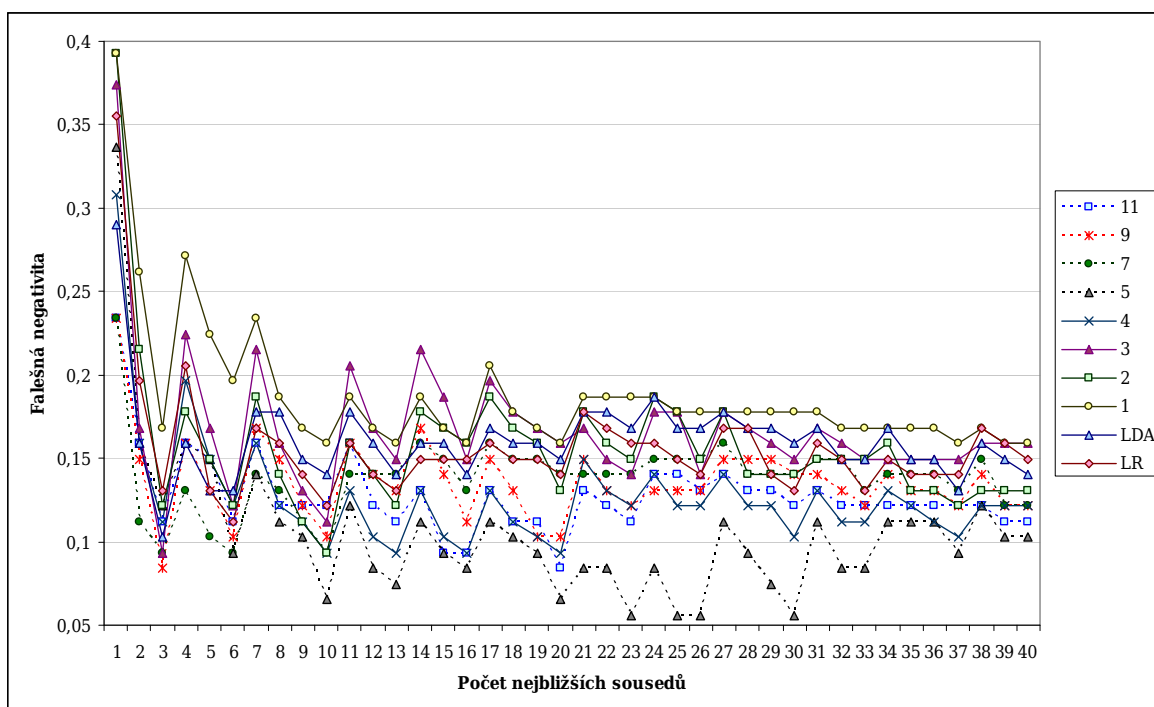
Zvyšování počtu hodnocených nejbližších sousedů na řešeném vzorku má tyto výsledky. U všech modelů strmě klesá falešná negativita při zvýšení  $k$  na 2 a 3. Vývoj falešné pozitivy je pro zvýšení  $k$  opačný - falešná pozitivita roste. Jak  $FPr$ , tak  $FNr$  s růstem  $k$  oscilují, na základě posouzení grafu lze usuzovat jejich určitou stabilizaci. Dále je možné zaznamenat, že nejjednodušší modely (1, 2 a 3 ukazatele) dávají horší výsledky (vyšší chybu) než ostatní. Obrázek 19 představuje vývoj celkové chyby, která je dána váženým průměrem  $FNr$  a  $FPr$ . S růstem  $k$  se oscilace celkové chyby stabilizují. Znatelně nejvyšší jsou celkové chyby u modelů „1“, „2“ a „3“. Minimální chybu dává model čtyř ukazatelů vybraných na základě krokové LDA ( $CF/A_1$ ,  $DR_1$ ,  $Zas/V_1$  a  $L2_1$ ) pro  $k = 17$ , jeho celková chyba je 12,5 %.

Obrázek 19 Metoda nejbližších sousedů - vývoj celkové chyby



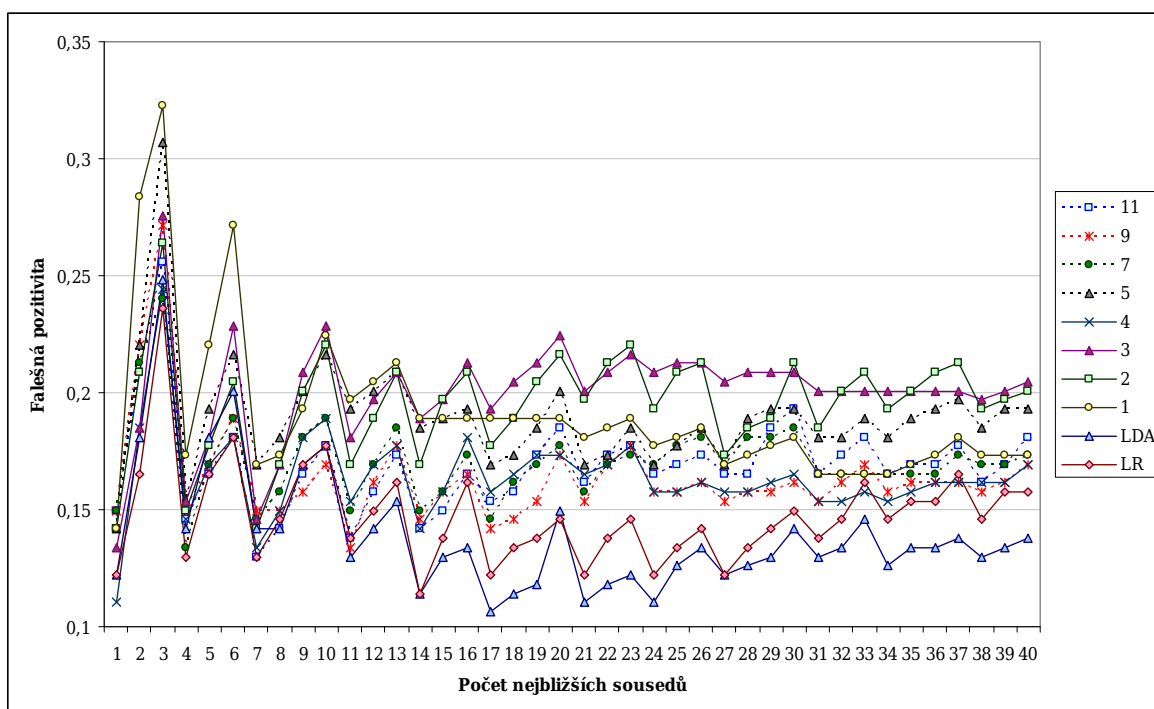
Pramen: vlastní zpracování

Obrázek 20 Metoda nejbližších sousedů - vývoj falešné negativity



Pramen: vlastní zpracování

Obrázek 21 Metoda nejbližších sousedů - vývoj falešné pozitivity



Pramen: vlastní zpracování

### Dílčí shrnutí

Metodou nejbližšího souseda ( $k = 1$ ) byly ověřovány klasifikační schopnosti třinácti modelů. Celková spolehlivost je nejvyšší u modelu, ve kterém jsou

zahrnuty ukazatele  $CF/A_1$ ,  $ZD/Mz_1$ ,  $CF/CZ_1$  a  $DR_1$ , a to 83,1 %. Při aplikaci metody  $k$ -nejbližších sousedů byl uvažován počet nejbližších sousedů  $k \in \{1, 2, \dots, 40\}$ . Při zvyšování počtu nejbližších sousedů nejprve strmě klesá falešná negativita, dále jak falešná pozitivita, tak falešná negativita se zvyšujícím se počtem nejbližších sousedů oscilují. Modely, v nichž jsou zařazeny jeden až tři ukazatele, poskytují horší výsledky než ostatní. Minimální chybu dává model ukazatelů vybraných na základě krokové lineární diskriminační analýzy pro  $k = 17$ , kdy jeho celková chyba je 12,5 %.

## 6.10. Metoda nejbližších prototypů

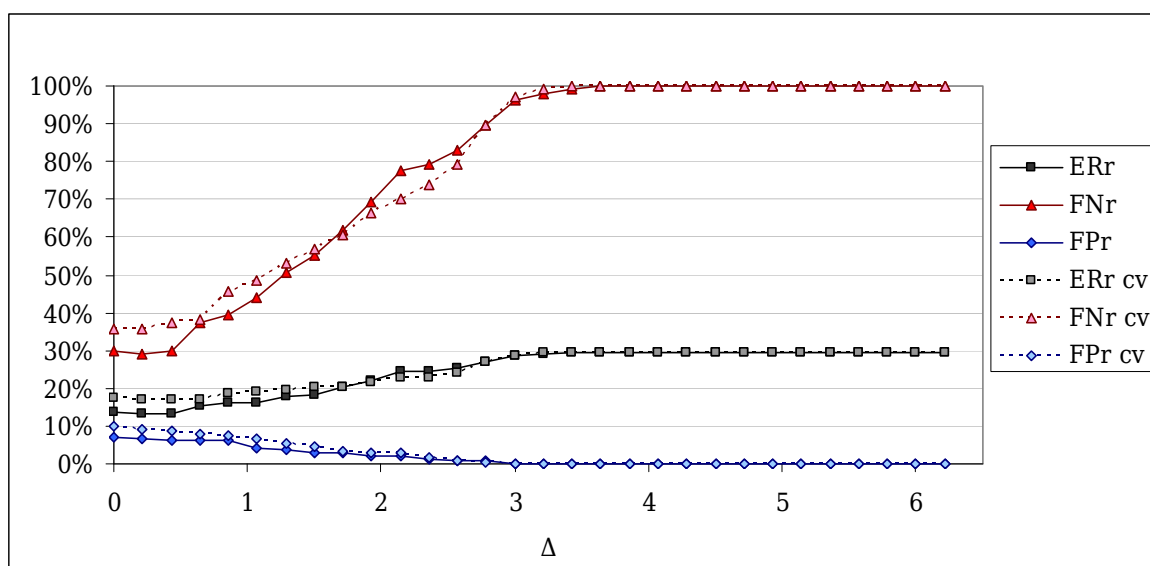
Vzhledem k tomu, že tato metoda byla navržena pro úlohy, kde počet znaků je mnohem vyšší než počet pozorování, je výchozí množinou všech 96 ukazatelů ( $32 \times 3$ , bez ukazatelů vztažených na pracovníka). Zde ani při zahrnutí všech znaků jejich počet nepřevyšuje počet pozorování ( $p = 96 < n = 361$ ), ale přesto ověříme chování a využitelnost této metody pro předpověď finanční tísně podniku. Při aplikaci jsou využity funkce z knihovny pamr (ve statistickém prostředí R).

Graf (Obrázek 22) zobrazuje vývoj celkové chyby, falešné negativy a falešné pozitivy při odhadu pomocí resubstituce a křížovým ověřováním. Pro  $\Delta = 0$  je do modelu zařazeno všech 96 ukazatelů a s rostoucí  $\Delta$  jejich počet klesá. Nejnížší celkové chyby dosahuje metoda nejbližších prototypů při  $\Delta = 0,214$  a při  $\Delta = 0,429$ , a to 13,3 % ( $FNr = 29$  % a  $FPr = 6,7$  % pro  $\Delta = 0,214$ ). Ovšem pro  $\Delta = 0,214$  obsahuje model 85 ukazatelů a pro  $\Delta = 0,429$  obsahuje 77 ukazatelů). Pro přijatelnou složitost modelu (do 10 ukazatelů) dosahuje falešná negativita 100 % a celková chyba odpovídá relativní četnosti ohrožených podniků.

### Dílčí shrnutí

Z výše uvedeného vyplývá, že aplikace metody nejbližších prototypů se ukázala jako nepříliš vhodná (tato metoda byla odvozena pro úlohy analýzy genů, kde počet znaků je obvykle mnohem větší než počet pozorování; Tibshirani et al. (2002) metodu odvodili pro klasifikační úlohu na vzorku, kde  $p \approx 2\,000$ ,  $n \approx 70$  a  $g = 4$ ). Lze se domnívat, že mimodiagonální prvky matic vnitroskupinové variability (které jsou v této metodě uvažovány nulové), nesou významnou informaci pro odvození klasifikátoru ohrožení finanční tísní.

Obrázek 22 Vývoj chyb - metoda nejbližších prototypů



Pramen: vlastní zpracování

## 6.11. Neuronové sítě

Tvorba klasifikačních modelů neuronových sítí byla provedena v programu Statistica, a to v módu automatické tvorby sítí. V tomto módu program sestavuje dle předem stanovených parametrů velký počet sítí, ze kterých jsou následně vybírány vhodné (podle dosažené spolehlivosti). Volitelnými parametry jsou typ sítě (vícevrstvá perceptronová síť nebo radial basis function), počet neuronů ve skryté vrstvě (dolní a horní mez), přenosová funkce ve skryté a výstupní vrstvě (identita, logistická, hyperbolický tangens či exponenciální) a chybová funkce (suma čtverců či křížová entropie). V módu vlastní tvorby sítí je možné navíc volit metodu trénovacího algoritmu, počet epoch či rychlost učení.

V případě neuronových sítí se vyskytuje několik specifíků, která budou dále stručně nastíněna. Datový soubor lze rozdělit nejen na skupinu trénovací a testovací, ale vytvořit i skupinu validační. Do učení se totiž zapojují pozorování ze skupiny jak trénovací, tak testovací; vlastnosti natrénované sítě jsou pak ověřeny na skupině validační. Z natrénovaných neuronových sítí je možné vytvářet jejich kolekce. Tyto kolekce kombinují přímo hodnoty výstupních neuronů jednotlivých sítí (ne až výsledné klasifikace). Analýzou citlivosti se v programu Statistica rozumí hodnocení významnosti vstupních neuronů. Pokud je citlivost (přesněji poměr citlivosti) rovna jedné, spolehlivost sítě se po jeho vyřazení nezmění; pokud je menší než jedna, jeho vyřazení by mělo zvýšit spolehlivost sítě.

Dále budou vytvářeny vícevrstvé perceptronové sítě s jednou skrytou vrstvou (dostupná varianta programu neumožňuje jinak). Relativní četnosti trénovací, testovací a validační množiny budou ponechány na implicitním nastavení programu Statistica, tedy trénovací 70 %, testovací 15 % a validační 15 %.

### 6.11.1. Neuronové sítě s 96 vstupními neurony

Nejprve byly sestavovány neuronové sítě s 96 vstupními neurony, tedy se všemi dostupnými ukazateli. Pomocí automatické tvorby sítí byly sestavovány sítě s 8 až 150 neurony ve skryté vrstvě, samotné trénování bylo časově poměrně náročné, proto bylo trénováno pouze 200 sítí.

Přes značně složitou architekturu sítě tyto (např. síť v konfiguraci 96-150-2 obsahuje téměř patnáct tisíc parametrů) dosahovaly spolehlivosti na validační skupině maximálně 88,6 % (na trénovací skupině až 100 %). Další postup nelze založit ani na analýze citlivosti, kde nejvyšší průměrné citlivosti dosáhl ukazatel  $Zas/V_1$  (1,14), u ostatních se průměrná citlivost pohybuje v intervalu  $\langle 0,99; 1,07 \rangle$ .

### 6.11.2. Neuronové sítě s 11 vstupními neurony

Vzhledem k neuspokojivým výsledkům příliš rozsáhlých neuronových sítí bude vstupy dalších neuronových sítí tvořit 11 výše vybraných ukazatelů (str. 88). V automatickém módu budou vytvářeny sítě s 5 až 20 neurony ve skryté vrstvě, přičemž bude vytvořeno celkem 100 000 sítí.

Natrénované sítě byly seřazeny podle spolehlivosti (1) na validační množině a (2) na testovací množině. Tabulka (Tab. 39) uvádí výsledky pěti nejspolehlivějších sítí včetně jejich aktivačních funkcí (ve skryté a výstupní vrstvě; exp. značí exponenciální funkci, logist. logistickou funkci, tanh hyperbolický tangens a softmax normovanou exponenciální funkci, u identity odpovídá výstup neuronu jeho vnitřnímu potenciálu). V jejich skrytých vrstvách je obsaženo 12 až 19 neuronů, pro další aplikaci těchto sítí je tedy nutno uchovat 170 až 268 parametrů (váhy jednotlivých spojení a prahové hodnoty).

Tab. 39 Přehled sítí s 11 vstupními neurony

Konfigurace	Spolehlivost (%)			Aktivační funkce	
	trénovací	testovací	validační	skrytá vrstva	výstupní vrstva
1. MLP 11-19-2	92,09	96,30	90,74	Exp.	Identita
2. MLP 11-12-2	92,09	94,44	90,74	Logist.	Softmax
3. MLP 11-14-2	91,70	92,59	90,74	Tanh	Softmax
4. MLP 11-20-2	91,30	94,44	90,74	Tanh	Softmax
5. MLP 11-13-2	90,91	92,59	90,74	Logist.	Softmax

Pramen: vlastní zpracování, STATISTICA 8

V tabulce (Tab. 39) je uvedena spolehlivost zvláště pro trénovací, testovací a validační množinu. V další tabulce (Tab. 40) jsou uvedeny chyby sítí na souboru

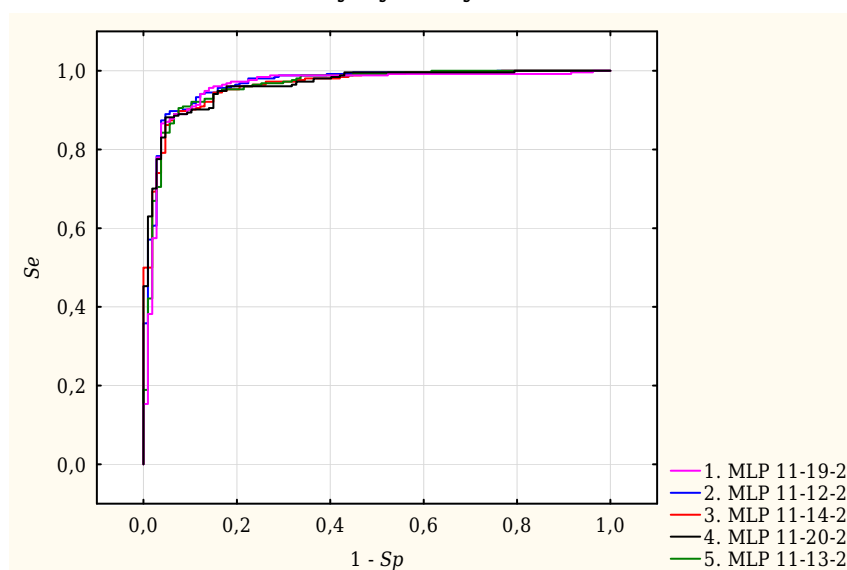
361 podniků a dále falešná negativita a falešná pozitivita. Celková chyba je nejnižší u sítě v konfiguraci 11-19-2, a to zejména vlivem nízké falešné positivity. Sít v konfiguraci 11-12-2 dosáhla velmi nízké falešné negativity, dokáže tedy s dobrou spolehlivostí ( $Se = 86,9\%$ ) odhalit ohrožené podniky. Na posledním řádku jsou uvedeny chyby kolekce těchto pěti sítí. Graf ROC křivek (Obrázek 23) dokládá výtečné klasifikační schopnosti těchto pěti neuronových sítí.

Tab. 40 Chyby sítí s 11 vstupními neurony

Konfigurace	$ERr$ (%)	$FNr$ (%)	$FPr$ (%)
1.MLP 11-19-2	7,48	16,82	3,54
2.MLP 11-12-2	7,76	13,08	5,51
3.MLP 11-14-2	8,31	14,95	5,51
4.MLP 11-20-2	8,31	15,89	5,12
5.MLP 11-13-2	8,86	15,89	5,91
Kolekce	8,59	15,89	5,51

Pramen: vlastní zpracování, STATISTICA 8

Obrázek 23 ROC křivky vybraných sítí 11 ukazatelů



Pramen: vlastní zpracování, STATISTICA 8

Pro další práci vyjdeme z výsledků analýzy citlivosti. U každé sítě bylo vytvořeno pořadí vstupů podle jejich citlivostního poměru (1. pořadí pro nejvyšší citlivostní poměr) a dále průměrné pořadí v těchto pěti sítích (Tab. 41). Na základě průměrného pořadí citlivostních poměrů je možné usuzovat na významnost jednotlivých vstupů, ukazatelů, a následně zjednodušit modely neuronových sítí zanedbáním nejméně významných vstupů.

Tab. 41 Analýza citlivosti sítí s 11 vstupními neurony

Ukazatel	1. MLP 11-19-2	2. MLP 11-12-2	3. MLP 11-14-2	4. MLP 11-20-2	5. MLP 11-13-2	Průměrné pořadí
$CF/A_1$	3,55	2,74	1,26	1,32	1,60	2,4
$Zas/V_1$	0,87	2,75	1,67	1,60	1,81	2,4
$CF/A_0$	1,00	2,24	1,35	1,21	1,51	3,8
$CF/CZ_1$	1,39	1,31	1,15	1,15	1,10	4,8
$DR_1$	0,83	1,45	1,33	1,31	1,29	5,0
$ZD/Mz_1$	12,02	1,59	1,06	1,04	1,02	5,8
$CF/CZ_{\delta'}$	0,49	1,29	1,11	1,09	1,06	7,2
$L2_1$	1,06	1,14	1,04	1,08	1,05	7,4
$Mt/V_1$	0,84	1,14	1,05	1,05	1,02	8,6
$DR_{\delta'}$	1,21	1,06	1,02	1,01	1,01	8,8
$ZD/W_{\delta'}$	1,15	1,00	1,00	1,00	1,00	9,8

Pramen: vlastní zpracování, STATISTICA 8

### 6.11.3. Neuronové sítě s 6 vstupními neurony

Tvorba neuronových sítí s 6 vstupními neurony vyplývá z výsledků předchozí části, konkrétně z analýzy citlivosti složitějších sítí. Na základě tabulky (Tab. 41) je vybráno 6 ukazatelů, jejichž vypuštění z modelu působí nejvyšší ztráty spolehlivosti. Jedná se o ukazatele  $CF/A_1$ ,  $Zas/V_1$ ,  $CF/A_0$ ,  $CF/CZ_1$ ,  $DR_1$ ,  $ZD/Mz_1$ . V automatickém módu tvorby sítí byly vytvářeny neuronové sítě s 4 až 12 neurony ve skryté vrstvě (celkem 100 000 sítí), z nichž 5 s nejvyšší spolehlivostí na validační, resp. na testovací množině uvádí tabulka (Tab. 42). Spolehlivost na validační množině je u některých sítí shodná se sítěmi složitějšími (Tab. 39). Nejlepšího výsledku dosáhla jedna z nejjednodušších sítí (v konfiguraci 6-4-2), tato síť obsahuje 38 parametrů, tedy zhruba  $\frac{1}{5}$  oproti modelům s 11 vstupy.

Tab. 42 Přehled sítí s 6 vstupními neurony

Konfigurace	Spolehlivost (%)			Aktivační funkce	
	trénovací	testovací	validační	skrytá vrstva	výstupní vrstva
1. MLP 6-4-2	95,26	94,44	90,74	Tanh	Softmax
2. MLP 6-4-2	95,26	88,89	90,74	Tanh	Softmax
3. MLP 6-11-2	87,75	87,04	88,89	Logist.	Logist.
4. MLP 6-10-2	87,75	87,04	88,89	Logist.	Logist.
5. MLP 6-6-2	87,75	87,04	88,89	Logist.	Softmax

Pramen: vlastní zpracování, STATISTICA 8



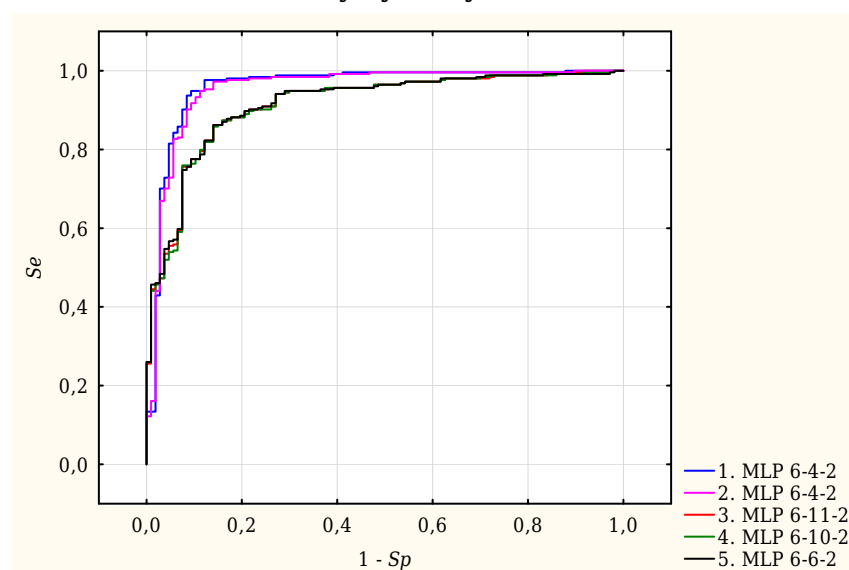
Výsledky ověření spolehlivosti, resp. chyby, sítí jsou uvedeny v tabulce (Tab. 43). Nejnižší chyby dosáhly nejjednodušší sítě (6-4-2), které mají velmi nízké hodnoty jak falešné pozitivivity, tak falešné negativity v porovnání nejen s ostatními sítěmi, ale i s výsledky metod aplikovaných v předchozích kapitolách. Nízká falešná negativita (vysoká senzitivita) je patrná i z grafu (Obrázek 24), kde ROC křivky těchto modelů jsou blízké spojnici bodů [0, 1] a [1, 1].

Tab. 43 Chyby sítí s 6 vstupními neurony

Konfigurace	<i>ERr</i> (%)	<i>FNr</i> (%)	<i>FPr</i> (%)
1. MLP 6-4-2	5,54	13,08	2,36
2. MLP 6-4-2	6,37	14,02	3,15
3. MLP 6-11-2	12,19	27,10	5,91
4. MLP 6-10-2	12,19	27,10	5,91
5. MLP 6-6-2	12,19	27,10	5,91
Kolekce	8,59	16,82	5,12

Pramen: vlastní zpracování, STATISTICA 8

Obrázek 24 ROC křivky vybraných sítí 6 ukazatelů



Pramen: vlastní zpracování, STATISTICA 8

Další zjednodušení architektury sítí může být opět založeno na výsledcích analýzy citlivosti (Tab. 44). Tři nejvýznamnější vstupy (ukazatele dlouhodobé rentability, vázanosti zásob a cash rentability aktiv) se shodují s výsledky krokové logistické regrese.

Tab. 44 Analýza citlivosti sítí s 6 vstupními neurony

Konfigurace		$DR_1$	$Zas/V_1$	$CF/A_1$	$CF/CZ_1$	$ZD/Mz_1$	$CF/A_0$
1. MLP 6-4-2	Citlivost	2,71	2,25	2,01	1,89	1,37	1,09
	Pořadí	1	2	3	4	5	6
2. MLP 6-4-2	Citlivost	2,35	2,61	2,24	1,75	1,53	1,13
	Pořadí	2	1	3	4	5	6
3. MLP 6-11-2	Citlivost	1,20	1,15	1,14	1,04	1,07	1,01
	Pořadí	1	2	3	5	4	6
4. MLP 6-10-2	Citlivost	1,20	1,14	1,14	1,04	1,07	1,01
	Pořadí	1	3	2	5	4	6
5. MLP 6-6-2	Citlivost	1,11	0,99	1,08	1,02	1,03	0,98
	Pořadí	1	5	2	4	3	6
Průměrné pořadí		1,2	2,6	2,6	4,4	4,2	6,0

Pramen: vlastní zpracování, STATISTICA 8

#### 6.11.4. Neuronové sítě s 3 vstupními neurony

Na základě výsledků analýzy citlivosti neuronových sítí s 6 vstupními neurony byly vybrány tři ukazatele s nejnižším průměrným pořadím poměru citlivosti, a to  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$ . I v tomto případě byla využita automatická tvorba sítí v programu Statistica, kde bylo vytvořeno 100 000 sítí s 2 až 6 neurony ve skryté vrstvě. Pět nejlepších sítí (podle spolehlivosti na validační, resp. testovací, resp. trénovací množině) včetně tvarů aktivačních funkcí je uvedeno v tabulce (Tab. 45). Výsledky analýzy citlivosti dávají jednoznačné výsledky, kde poměr citlivosti je u všech sítí nejvyšší u  $CF/A_1$  a nejnižší u  $Zas/V_1$  (ale vždy větší než jedna).

Tab. 45 Přehled sítí s 3 vstupními neurony

Konfigurace	Spolehlivost (%)			Aktivační funkce	
	trénovací	testovací	validační	skrytá vrstva	výstupní vrstva
1. MLP 3-2-2	88,93	87,04	94,44	Tanh	Softmax
2. MLP 3-4-2	88,93	87,04	94,44	Tanh	Softmax
3. MLP 3-6-2	88,54	87,04	94,44	Logist.	Softmax
4. MLP 3-5-2	88,54	87,04	94,44	Tanh	Logist.
5. MLP 3-3-2	88,54	87,04	94,44	Logist.	Softmax

Pramen: vlastní zpracování, STATISTICA 8

Tyto jednoduché sítě sice dosahují nižších spolehlivostí než složitější sítě natrénované v předchozích částech, jejich významnou vlastností je však vysoká spolehlivost na validační množině. Spolehlivost na validační množině je vyšší než na množině testovací a trénovací, což ukazuje na generalizační schopnosti těchto sítí (sít' tedy dokáže klasifikovat nová neznámá pozorování lépe než pozorování, na kterých byla natrénována).

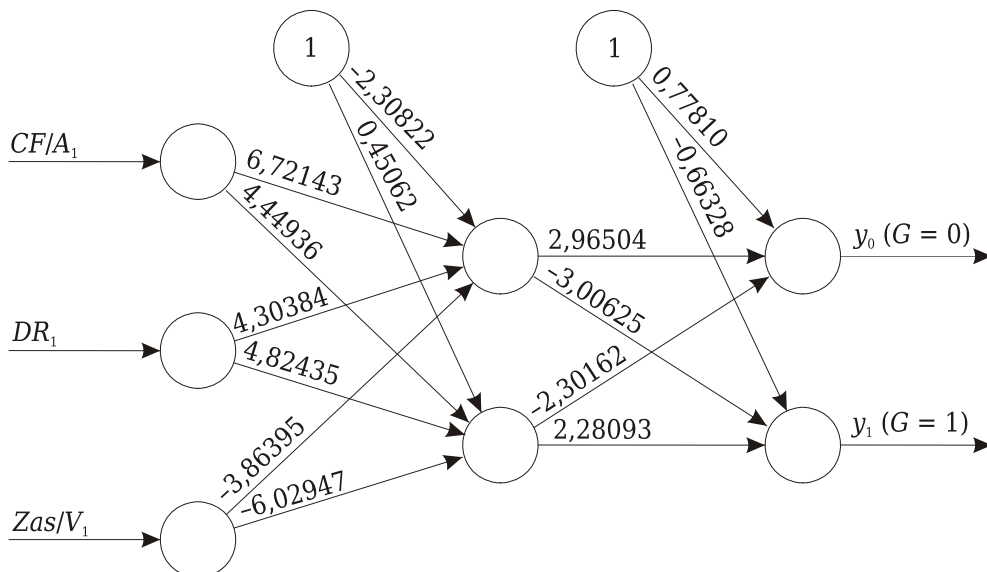
Tab. 46 Chyby sítí s 3 vstupními neurony

Konfigurace	<i>ERr</i> (%)	<i>FNr</i> (%)	<i>FPr</i> (%)
1. MLP 3-2-2	10,53	22,43	5,51
2. MLP 3-4-2	10,53	19,63	6,69
3. MLP 3-6-2	10,80	19,63	7,09
4. MLP 3-5-2	10,80	19,63	7,09
5. MLP 3-3-2	10,80	19,63	7,09
Kolekce	10,80	19,63	7,09

Pramen: vlastní zpracování, STATISTICA 8

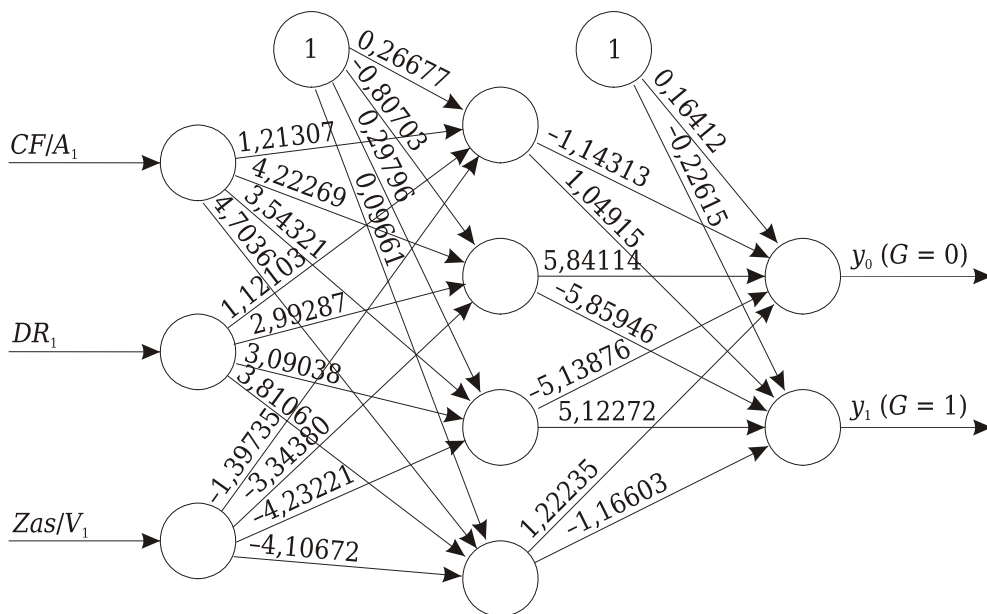
Architekturu jednodušších sítí je možné přehledně graficky znázornit včetně uvedení hodnot synaptických vah. V následujících grafech (Obrázek 25 a Obrázek 26) jsou zobrazeny topologie vybraných neuronových sítí - vícevrstevných perceptronových sítí v konfiguraci 3-2-2 a 3-4-2. Pro jejich další použití je třeba si uvědomit, že v programu Statistica jsou vstupní data normována, a to na rozpětí  $\langle 0, 1 \rangle$ . Hodnoty každého znaku u každého pozorování je tedy nutno normovat podle vztahu  $x^* = (x - x_{\min}) / R_x$ , kde  $x^*$  je normovaná hodnota znaku,  $x$  je naměřená hodnota znaku,  $x_{\min}$  je minimální hodnota znaku v souboru a  $R_x$  je rozpětí znaku  $x$ . Minimální a maximální hodnoty použitých znaků jsou uvedeny v tabulkách v příloze (Příloha B). V obou sítích je aktivační funkcí ve skryté vrstvě hyperbolický tangens a ve výstupní vrstvě softmax funkce (normovaná exponenciální funkce, Tab. 45).

Obrázek 25 Topologie neuronové sítě MLP 3-2-2



Pramen: vlastní zpracování

Obrázek 26 Topologie neuronové sítě MLP 3-4-2



Pramen: vlastní zpracování

### Dílčí shrnutí

Významné zjištění poskytuje srovnání spolehlivosti klasifikace složitějších a jednodušších neuronových sítí (z hlediska jejich architektury). V případě sítí s 96 vstupními neurony se projevila základní slabina vícevrstevných dopředných neuronových sítí, a to přeučení. Spolehlivost na validační skupině dosahovala maximálně 88,6 %, zatímco na skupině trénovací až 100 % - síť si tedy vzory „zapamatovala“ s nižší schopností zobecňování. Bylo by možné vytvořit síť, která si zapamatuje všechny vzory (síť dostatečné složitosti bez použití validační skupiny).

Dále tedy byly modely zjednodušovány, vstupní neurony nejprve tvořilo 11 dříve vybraných ukazatelů. V tomto případě se oproti složitým sítím zvýšila jak spolehlivost celková (až 92,5 %), tak spolehlivost na validační skupině (90,7 %). Přesto z analýzy citlivosti vytvořených sítí vyplývá u některých ukazatelů jejich nízká významnost, což poskytlo prostor pro další zjednodušení. Sítě s 6 vstupními neurony dosáhly vůbec nejvyšší celkové spolehlivosti (94,5 %) při spolehlivosti na validační skupině shodné se sítěmi s 11 vstupními neurony.

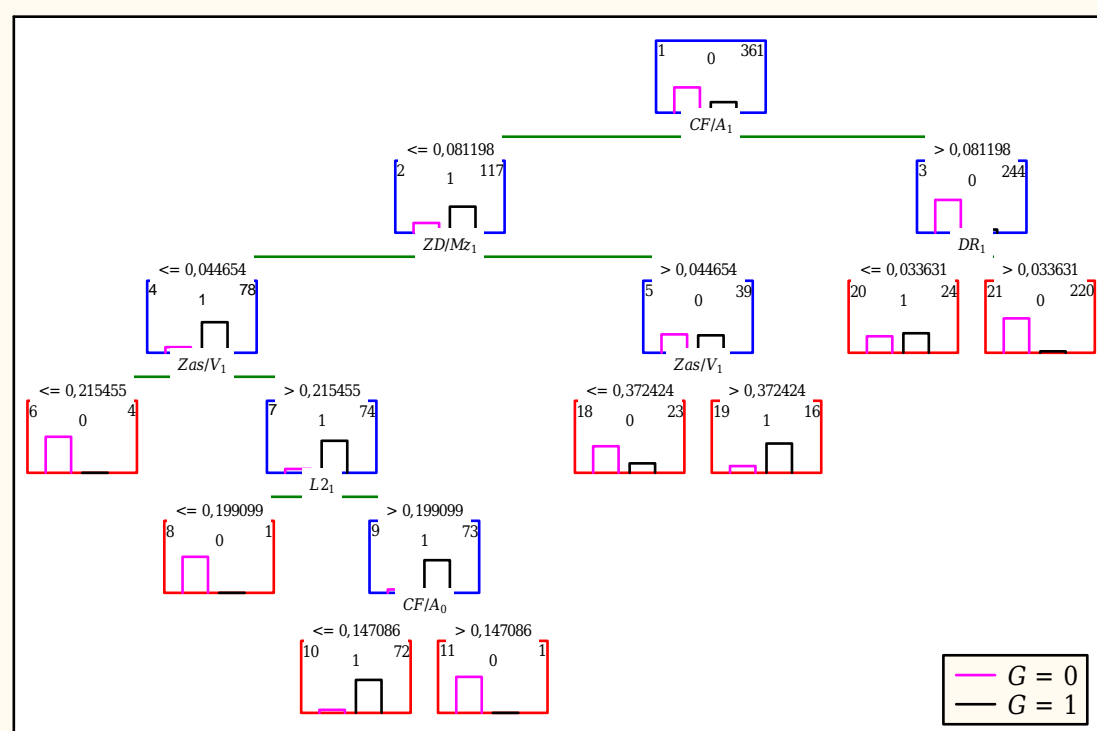
Opět na základě analýzy citlivosti byl model zjednodušen a dále byly konstruovány sítě pouze s ukazateli  $CF/A_1$ ,  $DR_1$  a  $Zas/V_1$ . Tyto jednoduché sítě projevily výtečné generalizační schopnosti, kdy spolehlivost pro neznámá pozorování dosáhla 94,4 % a přesahovala spolehlivost na trénovací a testovací skupině.

## **6.12. Klasifikační stromy a klasifikační lesy**

### **6.12.1. Klasifikační stromy**

Klasifikační stromy byly vytvářeny ve statistickém programu Statistica 8 pomocí metody C&RT (klasifikační a regresní stromy). Vstupní soubor byl opět tvořen z dříve vybraných 11 ukazatelů a na jejich základě byl vytvořen klasifikační strom (Obrázek 27). Podmínka pro ukončení dělení byla ponechána na doporučené hodnotě 36, tedy 10 % souboru (uzel, který obsahoval 36 či méně pozorování již nebyl dále dělen a byl označen za koncový) a pro výpočet nečistoty uzlů byla užita Giniho míra diverzity. Pod každým neterminálním uzlem je uveden znak, podle kterého se dělí a též jeho dělicí hodnota. V každém uzlu je uvedeno jeho číslo (levý horní roh), počet objektů v uzlu (pravý horní roh), objektům přiřazená třída (nahore uprostřed) a histogramy tříd (pro  $G = 1$  černě a pro  $G = 0$  růžově).

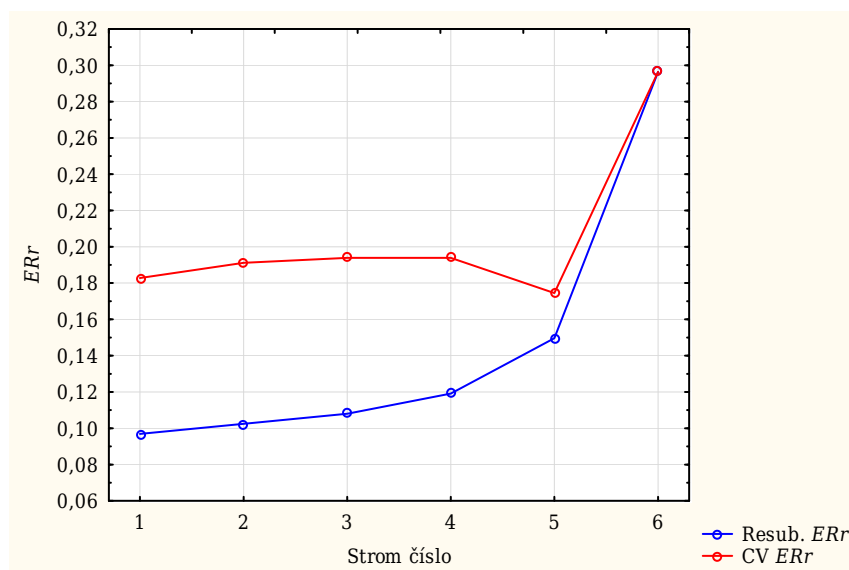
Obrázek 27 Neprořezaný klasifikační strom



Pramen: vlastní zpracování, STATISTICA 8

Výsledný strom obsahuje celkem 15 uzlů, z toho 8 koncových; pro dělení je užito 6 ukazatelů ( $CF/A_1$ ,  $ZD/Mz_1$ ,  $DR_1$ ,  $Zas/V_1$ ,  $L_{2_1}$  a  $CF/A_0$ ). Pomocí takto konstruovaného stromu je mylně klasifikováno 35 pozorování, a tedy  $ERr = 9,7\%$ ,  $FNr = 14\%$  a  $FPr = 7,9\%$ . Již z pohledu na strom můžeme usuzovat, že jeho prořezání je vhodné - některé listy mají nízké či dokonce jednotkové četnosti (uzly číslo 6, 8 a 11), některé listy jsou z pohledu na histogram poměrně „nečisté“ (četnost obou skupin je přibližně stejná, uzel číslo 20).

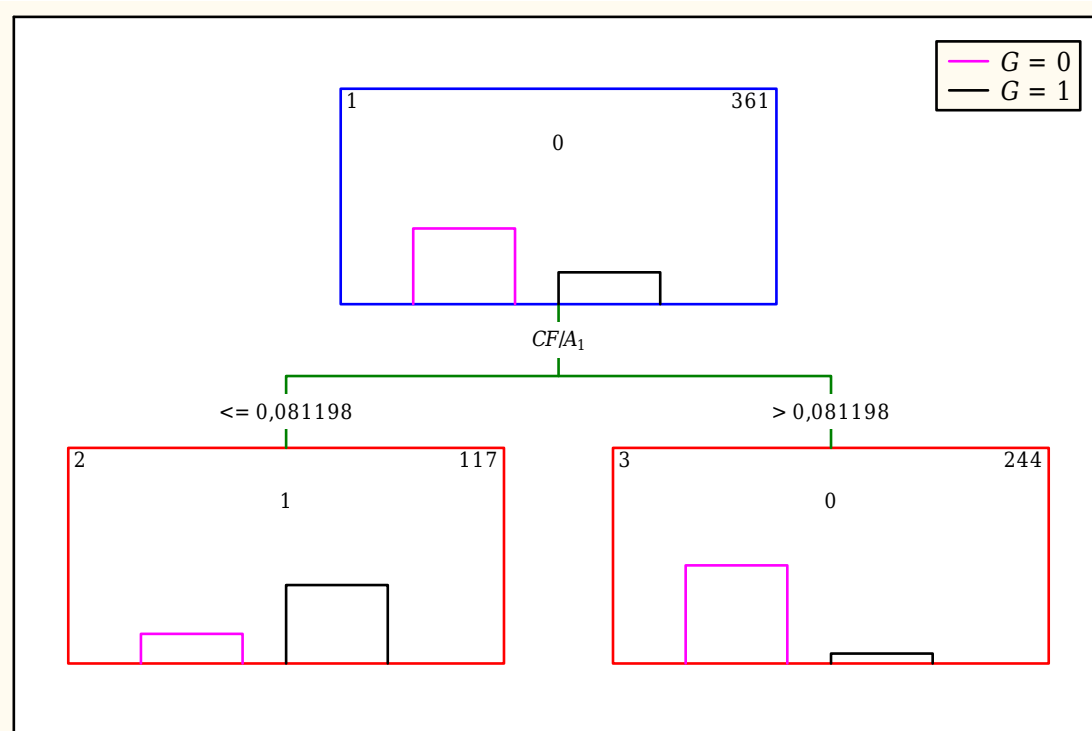
Obrázek 28 Proces prořezávání - vývoj celkové chyby



Pramen: vlastní zpracování, STATISTICA 8

Při prořezávání klasifikačního stromu je sledován vývoj celkové klasifikační chyby stromu, a to jak pomocí resubstituce, tak pomocí křížového ověřování (CV), v tomto případě je soubor rozdělen do 10 skupin (Obrázek 28). Při prořezávání (zjednodušování) stromu resubstituční chyba roste, CV chyba dosahuje minima u stromu č. 5 (17,5 %). Tento strom (Obrázek 29) má pouze dva koncové uzly a je uplatněno dělení podle ukazatele cash rentability aktiv. Chyba stromu ověřená pomocí resubstituce je 15 %,  $FNr = 20,6\%$  a  $FPr = 12,6\%$ . Strom č. 6 obsahuje pouze 1 uzel, tedy list, a v něm klasifikuje všechna pozorování do skupiny prosperujících podniků, celková chyba pak dosahuje relativní četnosti skupiny podniků ohrožených finanční tísňí.

Obrázek 29 Prořezaný klasifikační strom



Pramen: vlastní zpracování, STATISTICA 8

Pokud bychom změnili podmínku pro ukončení dělení, např. na 5 % pozorování, tedy 18, je výsledný strom mnohem košatější (obsahuje 37 uzlů, z toho 19 koncových). Ovšem po jeho prořezání získáme strom shodný s předchozím (Obrázek 29). Pokud jako míru nečistoty uzlů uijeme  $\chi^2$  (namísto Giniho míry diverzity), získáme totožný strom.

### 6.12.2. Klasifikační les

Lesy klasifikačních stromů budou konstruovány metodou Random Forest, kterou nabízí program Statistica. Jde tedy o sestavení velkého počtu klasifikačních stromů, na jejichž základě je postaveno klasifikační pravidlo. Vstupní datový

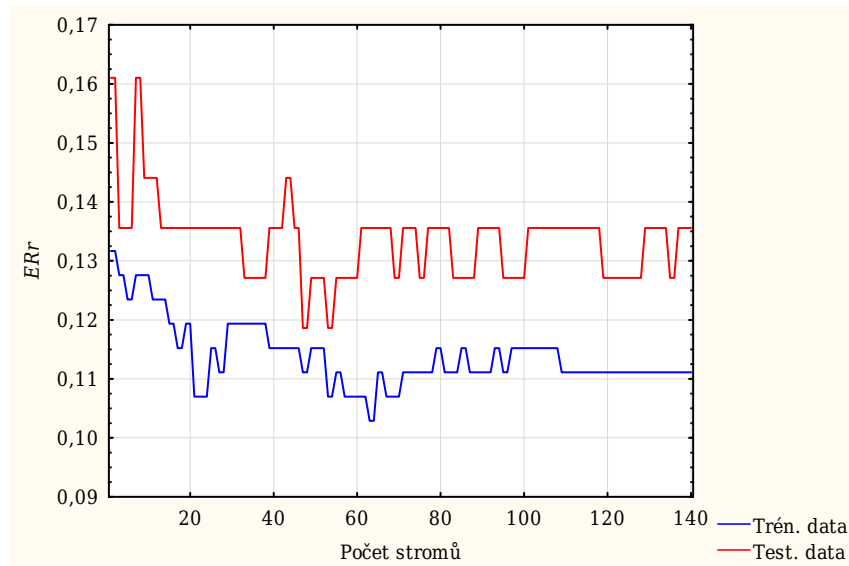
soubor je opět tvořen z 361 pozorování, u kterých je sledováno 11 výše vybraných ukazatelů.

Před tvorbou klasifikačního lesa je nutné nastavit několik vstupních parametrů, přičemž Statistica nabízí jejich doporučené hodnoty (na základě velikosti matice vstupních dat). Jedná se o počet prediktorů náhodně vybraných pro každý uzel, zde Statistica doporučuje hodnotu  $n = \log_2 M + 1$ , kde  $M$  je počet vstupních proměnných, v tomto případě tedy  $n = 4,46 \doteq 4$ . Rozsah testovacího souboru - ponecháno 30 %. Rozsah trénovacích souborů - ponecháno 50 % (jedná se o bootstrapové výběry). Počet stromů - zvýšeno na 1 000, přičemž tvorba může být předčasně ukončena (jak vyplývá z výsledků).

Dále je možné specifikovat podmínky pro dělení a omezit složitost stromů. Zde jsou opět ponechány doporučené hodnoty, tedy uzly obsahující 9 a méně pozorování se dále nedělí; pokud by dceřinný uzel obsahoval 5 či méně pozorování, dělení není provedeno; maximální počet úrovní stromů je 10 a maximální počet uzlů je 100 (je tedy umožněno konstruovat dosti složité - košaté - stromy).

Vývoj celkové chyby při konstrukci klasifikačního lesa uvádí Obrázek 30. Trénování je ukončeno po 140 krocích, tedy po konstrukci 140 stromů, kdy se chyba klasifikace na trénovací množině již nemění. Ukázkou čtyř zkonstruovaných stromů uvádí Obrázek 31, je zde patrné, že stromy se liší jak v ukazatelích použitých pro dělení, tak počtem uzlů - složitostí.

Obrázek 30 Vývoj celkové chyby klasifikačního lesa



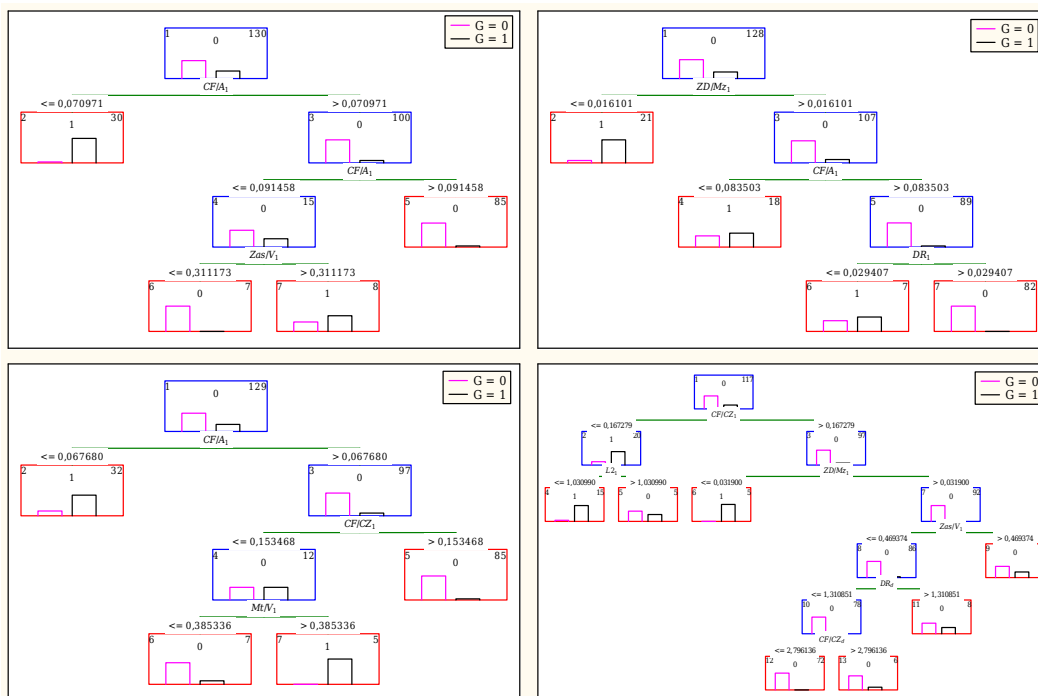
Pramen: vlastní zpracování, STATISTICA 8

Chyba na trénovacím souboru se ustálila na 11 %, chyba na testovacím souboru osciluje okolo 13 %. Celková chyba na celém souboru pak činí 12,2 %, falešná negativita je 22,4 % a falešná pozitivita 7,9 %. Kromě výše uvedených vstupních parametrů byly i ověřovány i jiné, výsledky (klasifikační chyba) byly však vždy horší (vyšší chyba na testovacím souboru, vyšší chyba na celkovém souboru, vyšší složitost - počet stromů; chyba na testovacím souboru byla vždy vyšší než na trénovacím souboru).



Pro každý z 11 ukazatelů je možné vyčíslit jeho významnost na základě vlivu ukazatele na pokles nečistoty uzlu (součtem přes všechny stromy a uzly). Hodnoty významnosti jsou normovány na rozpětí 0 až 100, nejvýznamnějším ukazatelem je  $CF/A_1$  (100), dále  $DR_1$  (88),  $ZD/Mz_1$  (85),  $CF/CZ_1$  (64),  $CF/A_0$  (62) a  $L2_1$  (61). Nejvýznamnější ukazatele  $CF/A_1$  a  $DR_1$  jsou obsaženy ve většině předchozích modelů, ale ukazatel  $Zas/V_1$ , který je také zahrnut ve většině předchozích modelů, je zde z hlediska významnosti až na 10. místě.

Obrázek 31 Ukázka stromů klasifikačního lesa



Pramen: vlastní zpracování, STATISTICA 8

## Dílčí shrnutí

Na základě 11 ukazatelů byl pomocí metody klasifikačních a regresních stromů vytvořen klasifikační strom s 8 listy, který dosahoval celkové spolehlivosti 90,3 %. Klasifikační strom byl dále prořezán, výsledný strom obsahuje pouze 2 listy, přičemž je uplatněno dělení pouze podle ukazatele cash rentability aktiv. U tohoto stromu byla při procesu prořezávání dosažena nejnižší chyba (17,5 %, křížové ověřování).

Lesy klasifikačních stromů byly konstruovány metodou Random Forest, přičemž tvorba lesa byla ukončena po konstrukci 140 klasifikačních stromů, kdy se chyba na testovacím souboru již nemění. Celková chyba tohoto lesa pak činí 12,2 %.

## 7. Závěr

Aplikační část disertační práce je založena na vytvořené definici podniku ohroženého finanční tísňí na základě relevantní literatury. Podnik je dle této definice považován za ohrožený finanční tísňí, pokud je záporný součet hospodářských výsledků za tři roky nebo je záporný peněžní tok (ze samofinancování) v kterémkoliv ze tří roků. Z hlediska aplikovaných klasifikačních metod se tedy jedná o vymezení závislé proměnné. Jako možné vysvětlující proměnné byla vybrána (s ohledem na jejich standardizaci) množina poměrových ukazatelů.

Práce je zpracována na výběrovém souboru českých zemědělských podniků, z čehož vyplývá možné omezení výsledků. Navazujícím krokem tedy může být ověřování přenositelnosti modelů, tedy jejich spolehlivosti na souborech podniků z jiných odvětví národního hospodářství.

Na konci března 2012 nebyl žádný ze 135 podniků zahrnutých ve vzorku uveden v Insolvenčním rejstříku. Skutečnost, že podnik mohl být v rejstříku uveden dočasně a posléze vymazán, nebyla ověřována a nelze ji vyloučit. Lze se tedy domnívat, že podniky obsažené ve vzorku patří do rozdílu množin podniků ohrožených finanční tísňí a podniků v úpadku (Obrázek 1, str. 13).

U modelů predikce finanční tísňě (včasného varování před finanční tísňí) je kromě nízké celkové chyby žádoucí jejich vysoká senzitivita (tedy nízká falešná negativita). Falešnou negativitu je v tomto kontextu možné přirovnat k pravděpodobnosti zaspání (nevarování před nastávajícími finančními problémy), naopak falešnou pozitivitu k pravděpodobnosti falešného poplachu. Zvyšování senzitivity je tedy žádoucí, ovšem při jejím zvyšování (nad optimální mez) se snižuje specifita a celkovou spolehlivost je nutno chápat právě jako kompromis (i ve formálním vyjádření se jedná o jejich vážený aritmetický průměr) senzitivity a specifity (respektive falešné negativity a falešné positivity).

### **Spolehlivost stávajících klasifikačních modelů**

Nejprve byla ověřována spolehlivost existujících klasifikačních modelů. Nejnižší celkové chyby dosáhl model IN95, Tafflerův model a Gurčíkův model. IN95 a Tafflerův model mají velmi vysokou specifitu (při senzitivě 45 %), naopak Gurčíkův model má velmi vysokou senzitivitu (94 % zároveň při dobré specifitě 65 %). Zde se projevují podmínky sestavení modelu, který je určen pro odhad budoucí prosperity slovenských zemědělských podniků.

### **Klasifikace pomocí jednorozměrných pravidel a jejich kombinací**

Do profilové analýzy bylo zařazeno 36 finančních poměrových ukazatelů, a to ze dvou období a též jejich relativní změna. Z prací zahraničních autorů vyplývá, že doplnění modelů o jiné než finanční ukazatele nepřináší užitek ve zvýšení jejich spolehlivosti. Přesto může být vhodné v budoucím zkoumání ověřovat při

predikci finanční tísně i klasifikační schopnosti nefinančních ukazatelů, s ohledem zkoumané odvětví např. výrobních ukazatelů.

Z výsledků profilové analýzy vyplynulo, že nejlépe lze rozdělit skupiny pomocí ukazatele cash rentability aktiv (se spolehlivostí 85,3 %, senzitivitou 68,2 % a specifitou 92,5 %). Klasifikační schopnosti tohoto ukazatele jsou velmi dobré a ukazatel je zahrnut do všech následně konstruovaných vícerozměrných modelů. Kombinace několika jednorozměrných klasifikátorů nepřinesly očekávaný efekt ve zvýšení spolehlivosti. Zvýšení spolehlivosti klasifikace, které přinášejí klasifikační pravidla založená na kombinaci ukazatelů, je statisticky nevýznamné (celková chyba 14,7 % u modelu ukazatele  $CF/A_1$  oproti 12,7 % modelu s devíti ukazateli) při současném nárůstu složitosti modelu. Snížení celkové chyby bylo způsobeno zvýšením specifity, ovšem na úkor senzitivity modelů - snížila se tedy spolehlivost včasného varování před blížícím se ohrožením finanční tísní.

### **Vícerozměrné klasifikační modely**

V tabulce (Tab. 47) jsou shrnuty dosažené celkové chyby, falešná negativita a falešná pozitivita vybraných dvaceti klasifikačních modelů. U uvedených modelů se celková chyba pohybuje od 5,5 % u neuronové sítě v konfiguraci 6-4-2 až po 34,4 % u kvadratické diskriminační funkce. Právě kvadratická diskriminační funkce dosahuje nejnižší falešné negativity, která je však kompenzována nejvyšší falešnou pozitivitou.

Při rozhodování o výběru modelu není vhodné se omezovat pouze na jejich spolehlivost (ověřovanou pomocí resubstituce), ale vzít v úvahu i další faktor, kterým je složitost modelu. Ten zahrnuje jak složitost výpočtu, tak rozsah uchovávaných parametrů (paměťové nároky). Složitost modelů pak výrazným způsobem ovlivňuje jejich případné praktické využití, např. v podnikové sféře. Z tohoto hlediska lze za nejjednodušší modely považovat jednorozměrný model založený za cash rentabilitě aktiv a prořezaný klasifikační strom (CT 5), který obsahuje pouze dva terminální a jeden neterminální uzel, tedy jedno klasifikační pravidlo. V těchto případech je složitost výpočtu i počet uchovávaných parametrů nízký.

Na druhé straně mezi nejsložitější modely patří les 140 klasifikačních stromů (CF 140), kde je počet uchovávaných parametrů velmi vysoký (ovšem při nízké složitosti dílčích výpočtů); a dále modely nejbližších sousedů. U nich je nezbytné uchovávat v paměti vybrané ukazatele pro celý soubor podniků (vzhledem k nutnosti práce s normovanými hodnotami ukazatelů i pravidlo pro normování nových pozorování), samotný výpočet vzdáleností není obtížný. Do skupiny složitějších modelů lze zařadit i kvadratické diskriminační funkce, u kterých je nutné pracovat (a samozřejmě je i uchovávat v paměti) s inverzními maticemi k skupinovým kovariančním maticím. U ostatních modelů jejich aplikace spočívá především v základních matematických operacích (u probitové regrese v případě dostupnosti tabelovaných hodnot distribuční funkce normálních normovaného rozdělení) a počet uchovávaných parametrů je rovněž nízký.

S ohledem na preferenci nízké falešné negativity je nutné zmínit i modely odvozené na základě krokové lineární diskriminační analýzy (LDF 4  $\theta$ ) a krokové logistické regrese (LR 3  $\theta$ ) s upraveným prahovým bodem (klasifikačním pravidlem). U těchto modelů právě úpravou prahové hranice klasifikačního pravidla došlo ke snížení falešné negativy při méně výrazném nárůstu falešné pozitivy, tím pádem se snížila i celková chyba (jak vyplývá ze srovnání chyb např. u LDF 4 a LDF 4  $\theta$ ).

Tab. 47 Shrnutí chyb vybraných modelů

Model	Celková chyba (%)	Falešná negativita (%)	Falešná pozitivita (%)
CF/A <sub>1</sub>	14,68	31,78	7,48
LDF 11	12,74	28,97	5,91
LDF 4	13,02	28,97	6,30
LDF 4 $\theta$	11,91	15,89	10,24
QDF 11	34,35	8,41	45,28
QDF 4	13,02	18,69	10,63
RLDF 4	12,47	29,91	5,12
LR 11	12,47	26,17	6,69
LR 3	13,02	25,23	7,87
LR 3 $\theta$	11,91	14,95	10,63
PR 11	13,02	28,97	6,30
PR 3	13,30	29,91	6,30
PR 3 $\theta$	12,19	20,56	8,66
1NN 4	16,90	30,84	11,02
17NN 4 LDA	12,47	16,82	10,63
MLP 11-19-2	7,48	16,82	3,54
MLP 6-4-2	5,54	13,08	2,36
MLP 3-4-2	10,53	19,63	6,69
CT 5	14,96	20,56	12,60
CF 140	12,19	22,43	7,87

Pramen: vlastní zpracování; CF/A<sub>1</sub> - jednorozměrný model pro cash rentabilitu aktiv, LDF 11 - lineární diskriminační funkce s 11 ukazateli; LDF 4 - lineární diskriminační funkce se 4 ukazateli, LDF 4  $\theta$  - lineární diskriminační funkce se 4 ukazateli s posunutou dělicí hodnotou; QDF 11 - kvadratická diskriminační funkce s 11 ukazateli; QDF 4 - kvadratická diskriminační funkce se 4 ukazateli; RLDF 4 - robustní lineární diskriminační funkce se 4 ukazateli; LR 11 - logistická regrese se 4 ukazateli; LR 3 - logistická regrese se 3 ukazateli; LR 3  $\theta$  - logistická regrese se 3 ukazateli s posunutou dělicí hodnotou; PR 11 - probitová regrese s 11 ukazateli; PR 3 - probitová regrese se 3 ukazateli; PR 3  $\theta$  - probitová regrese se 3 ukazateli s posunutou dělicí hodnotou; 1NN 4 - metoda nejbližšího souseda se 4 ukazateli; 17NN 4 LDA - metoda 17 nejbližších sousedů na základě ukazatelů vybraných pomocí krokové lineární diskriminační analýzy; MLP 11-19-2, MLP 6-4-2, MLP 3-4-2 - vícevrstvá dopředná neuronová síť v uvedené konfiguraci; CT 5 - klasifikační strom pro 5. kroku prořezávání; CF 140 - klasifikační les se 140 klasifikačními stromy.

Nejvyšší spolehlivosti tedy dosáhla neuronová síť v konfiguraci 6-4-2, ovšem tabulka (Tab. 47) neuvádí významné skutečnosti, které se týkají právě neuronových sítí. V případě nejjednodušších sítí (se třemi vstupními neurony, Tab. 45) je jejich spolehlivost na validační skupině vyšší než na skupině trénovací a testovací. Spolehlivost klasifikace pro nová neznámá pozorování je tedy vyšší než pro pozorování, na kterých byla síť natrénována. V tomto případě lze předpokládat, že se podařilo nalézt optimální konfigurace, které jsou dostatečně bohaté pro řešení klasifikační úlohy a zároveň dokáží zobecnit vztahy mezi vstupy (poměrovými ukazateli) a výstupem (tedy jevem finanční tíseň).

U některých pozorování je budoucí finanční tíseň či prosperita nepredikovatelným jevem. Ve výběrovém souboru se nachází jeden podnik (ze skupiny prosperujících), který byl pomocí všech modelů (uvedených v předchozí tabulce) klasifikován do skupiny podniků ohrožených finanční tísní. Z 10 podniků, které byly správně klasifikovány pomocí pouze jednoho modelu, bylo 5 prosperujících (správná klasifikace pouze některou z neuronových sítí) a 5 ohrožených finanční tísní (správná klasifikace pouze modelem QDA 11). Na druhou stranu 83,4 % pozorování bylo správně klasifikováno pomocí alespoň 16 modelů a 88,4 % pozorování bylo korektně zařazeno pomocí alespoň 10 modelů. Z toho vyplývá, že vytvoření kombinovaného modelu, jehož výsledek by byl dán hlasováním dílčích modelů by nepřineslo výsledek ve zvýšení spolehlivosti (nehledě na složitost takového modelu).

## Souhrn

Cílem disertační práce je ověřit možnosti vícerozměrných klasifikačních metod při predikci finanční tísně podniků v odvětví zemědělství. Aplikační část práce byla založena na definici podniku ohroženého finanční tísni vytvořené na základě relevantní literatury. Pokud byl součet hospodářských výsledků za tři roky záporný nebo cash flow bylo v kterémkoliv ze tří roků záporné, pak byl podnik označen jako ohrožený finanční tísni. V rámci řešení byla nejprve ověřena spolehlivost existujících klasifikačních modelů. Dále byly posouzeny schopnosti jednotlivých ukazatelů a jejich kombinací z hlediska spolehlivosti klasifikace. Hlavní část práce spočívala v konstrukci modelů pomocí vícerozměrných klasifikačních metod (diskriminační analýza lineární a kvadratická a jejich robustní varianty, metoda nejbližších sousedů a nejbližších prototypů, logistická regrese, probitová regrese, vícevrstvá perceptronová síť, klasifikační stromy a lesy).

Při ověřování spolehlivosti existujících klasifikačních modelů dosáhl nejnižší celkové chyby Tafflerův model (19,4 %), model IN95 (19,7 %) a Gurčíkův model (26 %). IN95 a Tafflerův model mají velmi vysokou specifitu, naopak Gurčíkův model má velmi vysokou senzitivitu (94 % při specifitě 65 %). Z výsledků profilové analýzy vyplynulo, že nejlépe lze rozdělit skupiny pomocí ukazatele cash rentability aktiv (s chybou 14,7 %). Kombinace několika jednorozměrných klasifikátorů nepřinesla očekávaný efekt ve zvýšení spolehlivosti.

U modelů vytvořených pomocí vícerozměrných klasifikačních metod se celková chyba pohybuje od 5,5 % u neuronové sítě v konfiguraci 6-4-2 až po 34,4 % u kvadratické diskriminační funkce. U modelů odvozených na základě lineární diskriminační analýzy a logistické regrese úpravou prahové hranice klasifikačního pravidla došlo ke snížení falešné negativy při méně výrazném nárůstu falešné positivity, tím pádem se snížila i celková chyba. Spolehlivost klasifikace nejjednodušších neuronových sítí se třemi vstupními neurony pro nová neznámá pozorování byla vyšší než pro pozorování, na kterých byla síť natrénována. Lze tedy předpokládat, že se podařilo nalézt optimální konfigurace, které jsou dostatečně bohaté pro řešení klasifikační úlohy a zároveň dokáží zobecnit vztahy mezi poměrovými ukazateli a jevem finanční tísně.

Klíčová slova: finanční tíseň, klasifikační metody, klasifikační modely

JEL: G30, C35

## Summary

The aim of the doctoral thesis is to screen possibilities of multivariate classification methods used for the prediction of a financial distress of agricultural enterprises. Application of the thesis was based on a definition of an enterprise threatened by financial distress defined according to relevant literature review. If the sum of the financial results for three years is negative or there was a negative cash flow in any of the three years the enterprise will be defined as threatened by financial distress. The reliability of current classification models was verified first as a part of the solution process. The ability of each indicator and their combinations in terms of reliability classification were assessed as well. The main part consisted in the construction of models using classification methods (linear and quadratic discriminant analysis and robust variants, the methods of nearest neighbours and prototypes, logistic regression, probit regression, multilayer perceptron networks, classification trees and forests).

When verifying the reliability of the existing classification models the lowest overall error was achieved by the Taffler's model (19.4%), followed by the IN95 model (19.7%) and by the Gurčík's model (26%). The IN95 and the Taffler model had very high specificity, while the Gurčík model had very high sensitivity (94% at specificity of 65%). The profile analysis revealed that it is the best to divide groups according to the cash return ratio (with the error of 14.7%). The combination of a few linear classifiers did not bring any expected effect of increased reliability.

For models created using multivariate classification methods, the total error varies from 5.5% for neural networks in the configuration of 6-4-2 to 34.4% for the quadratic discriminant function. For models based on linear discriminant analysis and logistic regression adjusting threshold limits of the classification rules reduced the false negative when less significant increase in false positives occurred, thus reducing the overall error. Reliability of classification of the simplest neural network with three input neurons for new unknown observations was greater than for the observations from the train sample. A success in finding the optimal configurations sufficiently rich to solve classification tasks and also able to generalize the relation between ratios and the phenomenon of the financial distress is assumed.

Key-words: financial distress, classification methods, classification models

JEL: G30, C35

## Literatura

- [1] ABID, F., ZOUARI, A. Financial distress prediction using neural networks: the Tunisian firms experience. *Modesfi Working Paper*, 2000. [online]. [cit. 3. 4. 2009]. Available at [www: http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=355980](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=355980).
- [2] ABIDALI, A. F., HARRIS, F. a methodology for predicting company failure in the construction industry. *Construction Management and Economics*, 1995, Vol. 13, pp. 189-196. ISSN 0144-6193.
- [3] ABOU EL SOOD, H. The usefulness of a composite model to failure prediction. In: *ABR & TLC Conference Proceedings*. Orlando, Florida, USA, 2008.
- [4] AGARWAL, V., TAFFLER, R. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance*, 2008, Vol. 32, pp. 1541-1551. ISSN 0378-4266.
- [5] AGARWAL, V., TAFFLER, R. J. Twenty-five years of the Taffler z-score model: does it really have predictive ability? *Accounting and Business Research*, 2007, Vol. 37 (4), pp. 285-300. ISSN 0001-4788.
- [6] AJVAZJAN, S., BEŽAJEVOVÁ, Z., STAROVEROV, O. *Metody vícerozměrné analýzy*. Praha: SNTL, 1981.
- [7] ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 1968, Vol. 23 (4), pp. 589-609. ISSN 0022-1082.
- [8] ALTMAN, E. I. Predicting Financial Distress of Companies: Revisiting the Z-score and ZETA<sup>®</sup> Models. In: *Bankruptcy, Credit Risk, and High Yield Junk Bonds*. 1st ed. Malden: Blackwell Publishers, 2002. ISBN 0-631-22563-3.
- [9] ALTMAN, E. I., GOODMAN, L. S. An economic and statistical analysis of the failing company. In: *Bankruptcy, Credit Risk, and High Yield Junk Bonds*, 1st ed. Malden: Blackwell Publishers, 2002. ISBN 0-631-22563-3.
- [10] ALTMAN, E. I., HALDEMAN, R. G., NARAYANAN, P. Zeta<sup>™</sup> analysis - a new model to identify bankruptcy risk of corporation. *Journal of Banking and Finance*, 1977, Vol. 1, pp. 29-54. ISSN 0378-4266.
- [11] ALTMAN, E. I., MARCO, G., VARETTO, F. Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance*, 1994, Vol. 18. pp. 505-529. ISSN 0378-4266.
- [12] ALTMAN, E. I., NARAYANAN, P. Business failure classification models: An international survey. In: *Bankruptcy, Credit Risk, and High Yield Junk Bonds*, 1st ed. Malden: Blackwell Publishers, 2002. ISBN 0-631-22563-3.
- [13] BAKER, H. K., POWELL, G. E. *Understanding Financial Management*. Malden: Blackwell Publishing, 2005. ISBN 0-631-23100-5.



- [14] BALCAEN, S., OOGHE, H. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 2006, Vol. 38, pp. 63-93. ISSN 0890-8389.
- [15] BEAVER, W. H. Financial ratios as predictors of failure. *Journal of Accounting Research*, 1966, Vol. 4, pp. 71-111. ISSN 0021-8456.
- [16] BECCHETTI, L., SIERRA, J. Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking and Finance*, 2003, Vol. 27, pp. 2099-2120. ISSN 0378-4266.
- [17] BEČVÁŘOVÁ, V. *Zemědělská politika*. Brno: MZLU, 2001. ISBN 80-7157-514-3.
- [18] BETINEC, M., PRCHAL, L. Poznámky k ROC křivkám. In: *ROBUST 2006. Sborník prací 14. zimní školy JČMF*. Praha: JČMF, 2006. ISBN 80-7015-073-4.
- [19] BISKUP, R. *Možnosti neuronových sítí*. Disertační práce. Praha: Česká zemědělská univerzita v Praze, Provozně ekonomická fakulta, 2009.
- [20] BLAHA, Z. S., JINDŘICHOVSKÁ, I. *Jak posoudit finanční zdraví firmy*, 3. vyd. Praha: Management Press, 2006. ISBN 80-7261-145-3.
- [21] BLUM, M. Failing company discriminant analysis. *Journal of Accounting Research*, 1974, Vol. 12 (1), pp. 1-25. ISSN 0021-8456.
- [22] BORTLÍČEK, Z. *ROC křivky*. Diplomová práce. Brno: Masarykova univerzita v Brně, Přírodovědecká fakulta, 2008.
- [23] BREIMAN, L. Bagging Predictors. *Machine Learning*, 1996, Vol. 24, pp. 123-140. ISSN 0885-6125.
- [24] BÜRGER, J. *Fraus velký ekonomický slovník: anglicko-český, česko-anglický*. Plzeň: Fraus, 2007. ISBN 978-80-7238-639-0.
- [25] CCB, Tisková zpráva Czech Credit Bureau, 17/8/2011. [online]. [cit. 18. 2. 2012]. Available at [www: http://www.creditbureau.cz/Novinky/Novinky/Pages/Od-roku-1993-zbankrotovalo-v-%C4%8Cesku-skoro-19-tis%C3%ADc-firem.aspx](http://www.creditbureau.cz/Novinky/Novinky/Pages/Od-roku-1993-zbankrotovalo-v-%C4%8Cesku-skoro-19-tis%C3%ADc-firem.aspx).
- [26] CCB, Tisková zpráva Czech Credit Bureau, 4/1/2012. [online]. [cit. 18. 2. 2012]. Available at [www: http://www.creditbureau.cz/Novinky/Novinky/Pages/V-roce-2011-v-%C4%8CR-zbankrotovalo-2-413-firem.aspx](http://www.creditbureau.cz/Novinky/Novinky/Pages/V-roce-2011-v-%C4%8CR-zbankrotovalo-2-413-firem.aspx).
- [27] ČEKIA, Tisková zpráva, 21. 10. 2010. [online]. [cit. 22. 3. 2012]. Available at [www: http://www.cekia.cz/images/tiskovezpravy/tz101021.pdf](http://www.cekia.cz/images/tiskovezpravy/tz101021.pdf).
- [28] ČEKIA, Tisková zpráva, 27. 2. 2012. [online]. [cit. 22. 3. 2012]. Available at [www: http://www.cekia.cz/cz/tiskove-zpravy/346-tz120227](http://www.cekia.cz/cz/tiskove-zpravy/346-tz120227).
- [29] ČERMÁKOVÁ, A. Verifikace klasifikačního pravidla v diskriminační analýze. In: *Kvantitativní metody v ekonomii 2004*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, 2004. ISBN 80-7040-691-7.

- [30] DAMBOLENA, I. G., KHOURY, S. J. Ratio stability and corporate failure. *The Journal of Finance*, 1980, Vol. 35 (4), pp. 1017-1026. ISSN 0022-1082.
- [31] DEAKIN, E. B. A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, 1972, Vol. 10 (1), pp. 167-179. ISSN 0021-8456.
- [32] DECLERC, M., HEINS, B., VAN WYMEERSCH, C. The use of value added ratios in statistical failure prediction models: some evidence on Belgian annual accounts. *Cahiers Economiques de Bruxelles*, 1992, Vol. 135, pp. 353-378. ISSN 0008-0195.
- [33] DOUCHA, R. *Bilanční analýza*. Praha: Grada Publishing, 1995. ISBN 80-85623-89-7.
- [34] DOUCHA, R. *Finanční analýza podniku: Praktické aplikace*. Praha: VOX, 1996. ISBN 80-902111-2-7.
- [35] DOUMPOS, M., ZOPOUNIDIS, C. *Multicriteria decision aid classification methods*. Dordrecht: Kluwer Academic Publishers, 2002. ISBN 1-4020-0505-8.
- [36] DRÁBEK, O., SEIDL, P., TAUFER, I. Umělé neuronové sítě – základy teorie a aplikace (2). *CHEMagazín*, 2005, Vol. XV (6), pp. 10-12. ISSN 1210-7409.
- [37] DRÁBEK, O., SEIDL, P., TAUFER, I. Umělé neuronové sítě – základy teorie a aplikace (3). *CHEMagazín*, 2006a, Vol. XVI (1), pp. 12-14. ISSN 1210-7409.
- [38] DRÁBEK, O., SEIDL, P., TAUFER, I. Umělé neuronové sítě – základy teorie a aplikace (4). *CHEMagazín*, 2006b, Vol. XVI (2), pp. 33-36. ISSN 1210-7409.
- [39] DRÁBKOVÁ, Z., KOUŘILOVÁ, J. Finanční zdraví podniku ve světle praktického využití kreativního účetnictví v podmínkách ČR. *Acta Universitatis Bohemicae Meridionales, The Scientific Journal for Economics, Management and Trade*, 2008, Vol. XI (1), pp. 61-65. ISSN 1212-3285.
- [40] EDMISTER, R. O. An empirical test of financial ratio analysis for small business failure prediction. *The Journal of Financial and Quantitative Analysis*, 1972, Vol. 7 (2), pp. 1477-1493. ISSN 0022-1090.
- [41] FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, 2006, Vol. 27, pp. 861-874. ISSN 0167-8655.
- [42] FORBELSKÁ, M. Neparametrická diskriminační analýza. In: *ROBUST 2000. Sborník prací jedenácté letní školy JČMF*. Praha: JČMF, 2001. ISBN 80-7015-792-5.
- [43] GENTRY, J. A., NEWBOLD, P., WHITFORD, D. T. Classifying Bankrupt Firms with Funds Flow Components. *Journal of Accounting Research*, 1985, Vol. 23 (1), pp. 146-160. ISSN 0021-8456.

- [44] GIROUX, G. *Financial Analysis: A User Approach*. Hoboken: Wiley, 2003. ISBN 0-471-46712-X.
- [45] GRAHAM, A. *Corporate Credit Analysis*. Chicago: Fitzroo Dearborn Publishers, 2000. ISBN 1-888998-75-X.
- [46] GRICE, J. S., DUGAN, M. T. The limitations of bankruptcy prediction models: some cautions for the researcher. *Review of Quantitative Finance and Accounting*, 2001, Vol. 17, pp. 151-166. ISSN 0924-865X.
- [47] GRICE, J. S., INGRAM, R. W. Tests of the generalizability of Altman's bankruptcy prediction model. *Journal of Business Research*, 2001, Vol. 54, pp. 53-61. ISSN 0148-2963.
- [48] GRÜNWARD, R. *Analýza finanční důvěryhodnosti podniku*. Praha: Ekopress, 2001. ISBN 80-86119-47-5.
- [49] GRÜNWARD, R., HOLEČKOVÁ, J. *Finanční analýza a plánování podniku*. Praha: Nakladatelství VŠE, 2004. ISBN 80-245-0684-X.
- [50] GURČÍK, L. G-index - metóda predikcie finančného stavu poľnohospodárskych podnikov. *Agricultural Economics*, 2002, Vol. 48 (8), pp. 373-387. ISSN 0139-570X.
- [51] HANLEY, J. A., MCNEIL, B. J. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, 1982, Vol. 143 (1), pp. 29-36. ISSN 0033-8419.
- [52] HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, Second edition. Springer, 2009. ISBN 978-0387848570.
- [53] HAVRÁNEK, T., VORLÍČEK, J. Lineární diskriminační funkce. In: *ROBUST '80*. Praha: JČMF, 1980.
- [54] HEBÁK, P. et al. *Vícerozměrné statistické metody [1]*. Praha: Informatorium, 2004. ISBN 80-7333-025-3.
- [55] HEBÁK, P. et al. *Vícerozměrné statistické metody [3]*. Praha: Informatorium, 2007. ISBN 978-80-7333-001-9.
- [56] HOLEČKOVÁ, J. *Finanční analýza firmy*, 1. vyd. Praha: ASPI - Wolters Kluwer, 2008. ISBN 978-80-7357-392-8.
- [57] HORÁKOVÁ, E. *Robustní metody v diskriminační analýze*. Diplomová práce. Brno: Masarykova univerzita, Přírodovědecká fakulta, 2008.
- [58] HOUDEK, M., SVOBODA, T., PROCHÁZKA, T. Klasifikace podle nejbližších sousedů. 2001. [online]. [cit. 7. 1. 2012]. Available at [www: http://cmp.felk.cvut.cz/cmp/courses/recognition/zapis\\_prednasky/zapis\\_01/4/rpz4.pdf](http://cmp.felk.cvut.cz/cmp/courses/recognition/zapis_prednasky/zapis_01/4/rpz4.pdf).
- [59] HUBERT, M., VAN DRIESSEN, K. Fast and robust discriminant analysis. *Computational Statistics & Data Analysis*, 2004, Vol. 45, pp. 301-320. ISSN 0167-9473.

- [60] CHARITOU, A., NEOPHYTOU, E., CHARALAMBOUS, C. Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 2004, Vol. 13 (3), pp. 465-497. ISSN 0963-8180.
- [61] JAKUBÍK, P. Makroekonomický model kreditního rizika. In: *Zpráva o finanční stabilitě*. Praha: ČNB, 2005.
- [62] JAKUBÍK, P., TEPLÝ, P. Skóring jako indikátor finanční stability. In: *Zpráva o finanční stabilitě*. Praha: ČNB, 2007. ISBN 978-80-87225-02-8.
- [63] JINDŘICHOVSKÁ, I., BLAHA, Z. S. *Podnikové finance*, 1. vyd. Praha: Management Press, 2001. ISBN 80-7261-025-2.
- [64] JOBSON J. D. *Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods*. New York: Springer, 1992. ISBN 0-387-97804-6.
- [65] KLASCHKA, J., KOTRČ, E. Klasifikační a regresní lesy. In: *ROBUST 2004. Sborník prací 13. letní školy JČMF*. Praha: JČMF, 2004. ISBN 80-7015-972-3.
- [66] KNÁPKOVÁ, A., PAVELKOVÁ, D. *Finanční analýza. Komplexní průvodce s příklady*. Praha: GRADA Publishing, 2010. ISBN 987-80-247-3349-4.
- [67] KOPTA, D. Possibilities of financial health indicators used for prediction of future development of agricultural enterprises. *Agricultural economics-Zemedska ekonomika*, 2009, Vol. 55 (3), pp. 111-125. ISSN 0139-570X.
- [68] KRALICEK, P. *Základy finančního hospodaření*. Praha: Linde Praha, 1993. ISBN 80-85647-11-7.
- [69] LACHER, R. C., COATS, P. K., SHARMA, S. C. FANT, L. F. A neural network for classifying the financial health of firm. *European Journal of Operational Research*, 1995, Vol. 85, pp. 53-65. ISSN 0377-2217.
- [70] LUSSIER, R. N., CORMAN, J. A success vs. failure prediction model of the manufacturing industry [online]. In: Conference of the Small Business Institute Director's Association, 1994. [cit. 16. 6. 2011]. Available at www: <http://sbaer.uca.edu/research/sbida/1994/pdf/48.pdf>.
- [71] MAREK, P. a kol. *Studijní průvodce financemi podniku*, 1. vyd. Praha: Ekopress, 2006. ISBN 80-86119-37-8.
- [72] MCLEAY, S., OMAR, A. The sensitivity of prediction models to the non-normality of bounded and unbounded financial ratios. *The British Accounting Review*, 2000, Vol. 32 (2), pp. 213-230. ISSN 0890-8389.
- [73] MELOUN, M., MILITKÝ, J. *Kompendium statistického zpracování dat: metody a řešené úlohy*, vyd. 2., přeprac. a rozš. Praha: Academia, 2006. ISBN 80-200-1396-2.
- [74] MELOUN, M., MILITKÝ, J. *Statistická analýza experimentálních dat*. Praha: Academia, 2004. ISBN 80-200-1254-0.
- [75] Metodika výpočtu finančního zdraví, [online]. [cit. 2. 1. 2012]. Available at www:

[http://www.szif.cz/irj/portal/anonymous/CmDocument?rid=%2Fapa\\_anon%2Fcs%2Fdokumenty\\_ke\\_stazeni%2Fefafd%2Fekonomika%2F1298035391331.pdf](http://www.szif.cz/irj/portal/anonymous/CmDocument?rid=%2Fapa_anon%2Fcs%2Fdokumenty_ke_stazeni%2Fefafd%2Fekonomika%2F1298035391331.pdf).

- [76] MICHA, B. Analysis of business failures in France. *Journal of Banking and Finance*, 1984, Vol. 8, pp. 281-291. ISSN 0378-4266.
- [77] MOSSMAN, C. E., BELL, G. G., SWARTZ, L. M., TURTLE, H. An empirical comparison of bankruptcy models. *The Financial Review*, 1998, Vol. 33, pp. 35-54. ISSN 0732-8516.
- [78] NEUMAIEROVÁ, I., NEUMAIER, I. Index IN - Index důvěryhodnosti českého podniku. *Terno*, 1995, Vol. 1 (5), pp. 7-10.
- [79] NEUMAIEROVÁ, I., NEUMAIER, I. Index IN05. In: *Sborník příspěvků z mezinárodní vědecké konference Evropské finanční systémy*. Brno: Masarykova univerzita v Brně, 2005. ISBN 80-210-3753-9.
- [80] NEUMAIEROVÁ, I., NEUMAIER, I. Proč se ujal index IN a nikoli pyramidový systém ukazatelů INFA. *Ekonomika a management*, 2008, Vol. 4. ISSN 1802-8934.
- [81] NEUMAIEROVÁ, I., NEUMAIER, I. *Výkonnost a tržní hodnota firmy*. Praha: GRADA Publishing, 2002. ISBN 80-247-0125-1.
- [82] OHLSON, J. A. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 1980, Vol. 18 (1), pp. 109-131. ISSN 0021-8456.
- [83] OOGHE, H., BALCAEN, S. Are failure prediction models transferable from one country to another? An empirical study using Belgian financial statements. Vlerick Leuven Gent Management School, *Working Paper Series* 2002/3.
- [84] OOGHE, H., SPAENJERS, C. A note on performance measures for failure prediction models. Vlerick Leuven Gent Management School, *Working Paper Series* 2006/29.
- [85] OOGHE, H., SPAENJERS, C., VANDERMOERE, P. Business failure prediction: simple-intuitive models versus statistical models. *Working Papers of Faculty of Economics and Business Administration*, Ghent University 05/338, 2005.
- [86] ORAVA, J. *Jádrové odhady a binární data*. Bakalářská práce. Brno: Masarykova univerzita v Brně, Přírodovědecká fakulta, 2006.
- [87] ORAVA, J. *Volba vyhlazovacího parametru při jádrových odhadech hustoty*. Diplomová práce. Brno: Masarykova univerzita v Brně, Přírodovědecká fakulta, 2008.
- [88] PECÁKOVÁ, I. Logistická regrese s vícekategoriální vysvětlovanou proměnnou. *Acta Oeconomica Pragensia*, 2007, Vol. 15 (1), pp. 86-96. ISSN 0572-3043.

- [89] PEEL, M. J., PEEL, D. A. Some further empirical evidence on predicting private company failure. *Accounting and Business Research*, 1987, Vol. 18 (69), pp. 57-66. ISSN 0001-4788.
- [90] PLATT, H. D., PLATT, M. B. A note on the use of industry-relative ratios in bankruptcy prediction. *Journal of Banking and Finance*, 1991, Vol. 15, pp. 1183-1194. ISSN 0378-4266.
- [91] RAO, C. R. *Lineární metody statistické indukce a jejich aplikace*. Praha: Academia, 1978.
- [92] RENCHER, A. C. *Methods of Multivariate Analysis*, Second edition. Wiley-Interscience, 2002. ISBN 0-471-41889-7.
- [93] ROSOCHATECKÁ, E., BERVIDOVÁ, L., SŮVOVÁ, H., TOMŠÍK, K. *Ekonomika podniků*. Praha: ČZU v Praze, 1999. ISBN 80-213-0480-4.
- [94] ROSOCHATECKÁ, E., ŘEZBOVÁ, H. Methodical approach to evaluation of financial health of agricultural enterprises in relation to the Sector Operational Program. *Agricultural Economics*, 2004, Vol. 50 (3), pp. 110-115. ISSN 0139-570X.
- [95] ROST, M., ČERMÁKOVÁ, A. Identifikace socio-ekonomických faktorů ovlivňujících mobilitu občanů venkova v Jihočeském kraji. *Acta regionalia et environmentalica*, 2007, Vol. 4 (1), pp. 23-28. ISSN 1336-9253.
- [96] ROST, M., TLUSTÝ, P. Klasifikační stromy jako alternativa k diskriminační analýze. In: *Medzinárodné vedecké dni 2008*. Nitra: FEM SPU v Nitre, 2008, pp. 1186-1190. ISBN 978-80-552-0060-6.
- [97] ROUSSEEUW, P. J., VAN DRIESSEN, K. A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, 1999, Vol. 41 (3), pp. 212-223. ISSN 0040-1706.
- [98] RŮČKOVÁ, P. *Finanční analýza. Metody, ukazatele, využití v praxi*, 3. rozšířené vydání. Praha: GRADA Publishing, 2010. ISBN 978-80-247-3308-1.
- [99] RÝDL, T. Vliv úpadkového práva na finanční stabilitu. In: *Zpráva o finanční stabilitě*. Praha: ČNB, 2005.
- [100] RYCHLÝ, M. Klasifikace a predikce. [online]. [cit. 10. 1. 2012]. Available at [www: http://www.fit.vutbr.cz/~rychly/docs/classification-and-prediction/classification-and-prediction.pdf](http://www.fit.vutbr.cz/~rychly/docs/classification-and-prediction/classification-and-prediction.pdf).
- [101] ŘEHÁKOVÁ, B. Nebojte se logistické regrese. *Sociologický časopis*, 2000, Vol. 36 (4), pp. 475-492. ISSN 0038-0288.
- [102] ŘEZANKOVÁ, H., HÚSEK, D. Klasifikace v programových systémech pro analýzu dat. In: *ROBUST 2000. Sborník prací jedenácté letní školy JČMF*. Praha: JČMF, 2001. ISBN 80-7015-792-5.
- [103] ŘEZANKOVÁ, H., HÚSEK, D., SNÁŠEL, V. *Shluková analýza dat*, 2. rozš. vyd. Praha: Professional Publishing, 2009. ISBN 978-80-86946-81-8.
- [104] SEDLÁČEK, J. *Finanční analýza podniku*. Brno: Computer Press, 2009. ISBN 978-80-251-1830-6.

- [105] SEDLAČÍK, M. *Vybrané metody klasifikace a jejich aplikace*. Disertační práce. Masarykova univerzita v Brně, Přírodovědecká fakulta, 2006.
- [106] SCHLESINGER, P. *Klasifikační postupy pro analýzu biologických dat*. Diplomová práce. Praha: Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, 2004.
- [107] STŘELEČEK, F., ZDENĚK, R. Importance of objective and formal adequacy for the indicators of enterprise financial health. *Agricultural economics-Zemedelska ekonomika*, 2004, Vol. 50 (12), pp. 543-551. ISSN 0139-570X.
- [108] STŘELEČEK, F., ZDENĚK, R. Finanční zdraví podniku. In: *Sborník mezinárodní konference Finance a účetnictví ve vědě, výuce a praxi*. Zlín: UTB ve Zlíně, 2005. ISBN 80-7318-288-2.
- [109] STŘELEČEK, F., ZDENĚK, R. Dependence of financial health on the objective and formal adequacy of the indicators. In: *BICABR 2006 - Brno International Conference on Applied Business Research*. Brno: MZLU v Brně, 2006. ISBN 80-7157-985-8.
- [110] SŮVOVÁ, H. *Finanční analýza v řízení podniku, v bance a na počítači*. Praha: Bankovní institut, 2000. ISBN 80-7265-027-0.
- [111] SYNEK, M. a kol. *Manažerská ekonomika*, 1. vyd. Praha, Grada Publishing, 1996. ISBN 80-7169-211-5.
- [112] ŠÍMA, J., NERUDA, R. *Teoretické otázky neuronových sítí*. Praha: Matfyzpress, 1996. ISBN 80-85863-18-9.
- [113] ŠKARPA, J. Bonitní model pro diagnózu firemní kondice. *Ekonom*, 2001, No. 7, pp. 28-29. ISSN 1210-0714.
- [114] TAFFLER, R. J. Empirical models for the monitoring of UK corporations. *Journal of Banking and Finance*, 1984, Vol. 8, pp. 199-227. ISSN 0378-4266.
- [115] TAFFLER, R. J. Forecasting company failure in the UK using discriminant analysis and financial ratio data. *Journal of the Royal Statistical Society. Series A (General)*, 1982, Vol. 145 (3), pp. 342-358. ISSN 0035-9238.
- [116] TAUFER, I., DRÁBEK, O., SEIDL, P. Umělé neuronové sítě - základy teorie a aplikace (5). *CHEMagazín*, 2006c, Vol. XVI (5), pp. 29-31. ISSN 1210-7409.
- [117] TAUFER, I., DRÁBEK, O., SEIDL, P. Umělé neuronové sítě - základy teorie a aplikace (7). *CHEMagazín*, 2007, Vol. XVII (3), pp. 2-7. ISSN 1210-7409.
- [118] TIBSHIRANI, R., HASTIE, T., NARASIMHAN, B., CHU, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, Vol. 99 (10), pp. 6567-6572.
- [119] VALACH, J. a kol. *Finanční řízení podniku*, 2. vyd. Praha: Ekopress, 1999. ISBN 80-86111-21-1.

- [120] XU, X., WANG, Y. Financial failure prediction using efficiency as a predictor. *Expert Systems with Applications*, 2009, Vol. 36, pp. 366-373. ISSN 0957-4174.
- [121] YANG, Z.R., JAMES, H., PACKER, A. The failure prediction of UK private construction companies. RICS, London, COBRA 1997, 1997. [online]. [cit. 21. 2. 2011]. Available at [www: http://www.rics.org/site/download\\_feed.aspx?fileID=2726&fileExtension=PDF](http://www.rics.org/site/download_feed.aspx?fileID=2726&fileExtension=PDF).
- [122] Zákon č. 182/2006 Sb., o úpadku a způsobech jeho řešení.
- [123] ZMIJEWSKI, M. E. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 1984, Vol. 22, pp. 59-82. ISSN 0021-8456.
- [124] *Zpráva o stavu zemědělství ČR za rok 2009 „ZELENÁ ZPRÁVA“*. Praha: Ministerstvo zemědělství, 2010. ISBN 978-80-7084-940-8.



## Seznam obrázků

Obrázek 1 Průnik finanční tísně a úpadku	13
Obrázek 2 Schéma tvorby klasifikačního modelu	20
Obrázek 3 Schéma biologického neuronu	38
Obrázek 4 Formální neuron	38
Obrázek 5 Příklad klasifikačního lesa vytvořeného metodou bagging	46
Obrázek 6 Příklady průběhu ROC křivek	50
Obrázek 7 Trade-off funkce	51
Obrázek 8 Vývoj vybraných poměrových ukazatelů	54
Obrázek 9 Diskriminační skóre a skupinové centroidy	57
Obrázek 10 Průměrné směrodatné odchylky vs. doba před úpadkem	70
Obrázek 11 Vývoj chyb u ukazatele cash rentabilita aktiv	80
Obrázek 12 ROC křivka pro CZ/A	86
Obrázek 13 ROC křivky vybraných ukazatelů	87
Obrázek 14 ROC křivky diskriminačních funkcí	95
Obrázek 15 Wafer graf $CF/A_1$ , $DR_1$ a $Zas/V_1$	99
Obrázek 16 ROC křivky logistických modelů	100
Obrázek 17 ROC křivky probitových modelů	103
Obrázek 18 Metoda nejbližšího souseda - dosažená spolehlivost	105
Obrázek 19 Metoda nejbližších sousedů - vývoj celkové chyby	106
Obrázek 20 Metoda nejbližších sousedů - vývoj falešné negativy	107
Obrázek 21 Metoda nejbližších sousedů - vývoj falešné pozitivy	107
Obrázek 22 Vývoj chyb - metoda nejbližších prototypů	109
Obrázek 23 ROC křivky vybraných sítí 11 ukazatelů	111
Obrázek 24 ROC křivky vybraných sítí 6 ukazatelů	113
Obrázek 25 Topologie neuronové sítě MLP 3-2-2	116
Obrázek 26 Topologie neuronové sítě MLP 3-4-2	116
Obrázek 27 Neprořezaný klasifikační strom	118
Obrázek 28 Proces prořezávání - vývoj celkové chyby	118
Obrázek 29 Prořezaný klasifikační strom	119
Obrázek 30 Vývoj celkové chyby klasifikačního lesa	120
Obrázek 31 Ukázka stromů klasifikačního lesa	121

## Seznam tabulek

Tab. 1 Rozdělení podniků dle ČEKIA Stability Rating	8
Tab. 2 Klasifikační matice	48
Tab. 3 Váhy ukazatelů indexu IN95	55
Tab. 4 Kralickův Quicktest - ukazatele	61
Tab. 5 Kralickův Quicktest - hodnotící tabulka	61
Tab. 6 Bodové ohodnocení ukazatelů dle metodiky výpočtu finančního zdraví	63
Tab. 7 Definice kategorií dle metodiky výpočtu finančního zdraví	64
Tab. 8 Tamariho model	66
Tab. 9 Regresní koeficienty jednotlivých modelů	68
Tab. 10 Rozdělení souboru	77
Tab. 11 Vztah základního a výběrového souboru	77
Tab. 12 Spolehlivost existujících klasifikačních modelů	79
Tab. 13 Jednorozměrná klasifikační pravidla dle jejich spolehlivosti	81
Tab. 14 Matice spearmanových korelačních koeficientů	82
Tab. 15 Rozdělení podniků podle počtu správných klasifikací	82
Tab. 16 Kombinace třech ukazatelů	83
Tab. 17 Kombinace pěti ukazatelů	84
Tab. 18 Kombinace sedmi ukazatelů	84
Tab. 19 Kombinace devíti ukazatelů	85
Tab. 20 Plochy pod ROC křivkou a minimální vzdálenosti od bodu [0, 1]	86
Tab. 21 Matice spearmanových korelačních koeficientů	88
Tab. 22 Grubbsův test odlehlých pozorování	89
Tab. 23 Kolmogorovův-Smirnovův test normality	90
Tab. 24 Testy shody středních hodnot mezi $G = 0$ a $G = 1$	90
Tab. 25 Párové testy shody středních hodnot mezi $t = 0$ a $t = 1$	91
Tab. 26 Klasifikační matice - LDA se 4 ukazateli	93
Tab. 27 Klasifikační matice - QDA se 4 ukazateli	94
Tab. 28 Podíl chyb - robustní LDA, 11 ukazatelů	95
Tab. 29 Podíl chyb - robustní LDA, 4 ukazatele	96
Tab. 30 Podíl chyb - robustní QDA, 11 ukazatelů	96
Tab. 31 Podíl chyb - robustní QDA, 4 ukazatele	97
Tab. 32 Logistická regrese - odhady parametrů, 11 ukazatelů	98
Tab. 33 Logistická regrese - odhady parametrů, 3 ukazatele	99
Tab. 34 Klasifikační matice - logistická regrese, 3 ukazatele	99
Tab. 35 Probitová regrese - odhady parametrů, 11 ukazatelů	101
Tab. 36 Probitová regrese - odhady parametrů, 3 ukazatele	102
Tab. 37 Klasifikační matice - probitová regrese, 3 ukazatele	102
Tab. 38 Metoda nejbližšího souseda - dosažená spolehlivost	104

Tab. 39 Přehled sítí s 11 vstupními neurony	110
Tab. 40 Chyby sítí s 11 vstupními neurony	111
Tab. 41 Analýza citlivosti sítí s 11 vstupními neurony	112
Tab. 42 Přehled sítí s 6 vstupními neurony	112
Tab. 43 Chyby sítí s 6 vstupními neurony	113
Tab. 44 Analýza citlivosti sítí s 6 vstupními neurony	114
Tab. 45 Přehled sítí s 3 vstupními neurony	114
Tab. 46 Chyby sítí s 3 vstupními neurony	115
Tab. 47 Shrnutí chyb vybraných modelů	124

# Seznam příloh

Příloha A

Příloha B

# Příloha A

## Roztřídění podniků do skupin

Pokud byla u podniku (v prvním sloupci uvedeno identifikační číslo) dostupná data v pětileté časové řadě, jsou vyplněny údaje ve sloupci, který odpovídá období  $t = 0$ . První číslo v buňce pak je kumulovaný hospodářský výsledek z běžné činnosti za období  $t = 2, 3$  a  $4$ , další čísla odpovídají cash flow v letech  $t = 2, t = 3$  a  $t = 4$  (vše v tis. Kč).

Tab. A-1 Roztřídění podniků do skupin

Podnik	2000	2001	2002	2003	2004	2005	2006
105309				8204 14839 16374 14905	7552 16374 14905 12985		
107361							42346 31474 25955 24024
107956	8356 23623 17400 31588	22857 17400 31588 37552	42588 31588 37552 30511	49401 37552 30511 37912	37723 30511 37912 22370	25721 37912 22370 19547	15314 22370 19547 27863
107999			47157 44649 53497 32062	60813 53497 32062 63025	52345 32062 63025 51827	61175 63025 51827 52481	56693 51827 52481 67350
109207			22876 21532 13591 22515	23410 13591 22515 23581	21802 22515 23581 21709	14451 23581 21709 15970	
109568							1641 14528 5802 14088
109916				8743 18348 21373 26556	11182 21373 26556 30141	12149 26556 30141 25621	9475 30141 25621 23666
109959				9849 14850 9030 11435	930 9030 11435 9058		
109975				13784 19296 24868 30991	8856 24868 30991 17604		
110205		-8738 5000 10587 12030					
110515				23112 16007 22132 21096	21137 22132 21096 16274	6596 21096 16274 6809	6976 16274 6809 20662

Podnik	2000	2001	2002	2003	2004	2005	2006
110591							3041 9468 7201 6958
110663	6979 19597 11002 19371	8689 11002 19371 23175	12595 19371 23175 26663	17764 23175 26663 34652	14558 26663 34652 24941	4597 34652 24941 19621	91 24941 19621 25215
110680					3564 9526 9523 9379	3617 9523 9379 9083	3604 9379 9083 7266
112356							2384 14607 8354 11979
112437					-4273 10262 11849 1909	-14650 11849 1909 11804	
112623	-2710 4509 -4166 14740	-1448 -4166 14740 9457					
113085					9749 5946 11759 8883	9622 11759 8883 6609	4898 8883 6609 6136
114057	-4011 -1859 -1793 2373	-1099 -1793 2373 2690	1999 2373 2690 3535	2534 2690 3535 5793	91 3535 5793 2610	-5732 5793 2610 383	-9843 2610 383 1111
115088	-6237 61 -146 3941	-2837 -146 3941 3056					
115436	-6537 1005 3795 10518						
116084	-6913 3104 -2449 4507	-8616 -2449 4507 2693	-956 4507 2693 7983	-262 2693 7983 8647	-2401 7983 8647 1733		
116734				15738 21167 15707 16421	10527 15707 16421 14108	7347 16421 14108 18664	5019 14108 18664 13033
117099				-7221 2603 3778 2689	-8327 3778 2689 1976	-13607 2689 1976 -1315	
118605		-1508 -3627 8542 621					
124851			9473 11045 9944 10720				
129046					-121 3365 3162 3702	-3898 3162 3702 2676	

Podnik	2000	2001	2002	2003	2004	2005	2006
129488				121683 70770 61493 116994	151685 61493 116994 106283	137618 116994 106283 63047	97264 106283 63047 71875
131351					8871 9678 23260 10854	2343 23260 10854 3003	
137235						-2721 12538 16179 -826	-2006 16179 -826 22421
139076						30649 34990 37188 19950	19336 37188 19950 26310
139301				7905 14395 14665 17430	2870 14665 17430 11182		
143146				2047 70 8564 7878			
143308			8375 20512 11211 7841	12672 11211 7841 21057	26756 7841 21057 26494	19113 21057 26494 -1095	
144924							13859 34007 22283 28316
144991				2589 9066 5870 11258	-5794 5870 11258 -1245	-8500 11258 -1245 6583	
150657		11586 9864 14581 13645	12918 14581 13645 13217				
156400	-107 395 7863 12525						14228 15849 6780 8408
488682						-8360 1231 2363 5128	1099 2363 5128 11646
655961						1170 7049 6663 5301	2950 6663 5301 4377
13693484				7747 12569 10319 10389	7255 10319 10389 12513	3651 10389 12513 7480	1730 12513 7480 6125
18251978				12520 5303 4138 8867			
25173189					2311 9992 17746 14011	-1359 17746 14011 7642	

Podnik	2000	2001	2002	2003	2004	2005	2006
25179403						23354 33346 38830 31486	18046 38830 31486 23214
25183907	-1945 1843 4727 7650						
25183982	-23778 -1495 -5438 1921	-13876 -5438 1921 11038	-3121 1921 11038 6419				
25227432	-3246 1610 -1099 10384	380 -1099 10384 8213					
25250213				-3041 7834 7266 13140	-6750 7266 13140 6135	-6056 13140 6135 9388	
25256114		-3532 -1273 4027 387					
25258729		-5146 -4921 10593 917	-1040 10593 917 544				
25265521						-750 6479 6066 2110	
25279327				-2364 6539 4582 9325	-7376 4582 9325 2141	-11562 9325 2141 1998	
25297091				4953 5685 8417 6233			
25297228					-4400 6224 7683 5191	-12522 7683 5191 -3051	
25303589			8550 12347 12974 8475	9312 12974 8475 10542	12076 8475 10542 16171	12884 10542 16171 12247	10530 16171 12247 8579
25331850							2087 9279 9934 9577
25370596			3218 4721 1226 3599	2029 1226 3599 4225			
25398849			2140 6601 9876 8012				
25504932						-10301 14032 18553 -725	



Podnik	2000	2001	2002	2003	2004	2005	2006
25571095			44403 33565 36986 29136	55458 36986 29136 54733	95718 29136 54733 87776	87940 54733 87776 27192	67075 87776 27192 32515
25573004			6122 11304 8534 6126	9390 8534 6126 14733			
25582844		-1152 -640 3668 2519	-655 3668 2519 1921	2379 2519 1921 8821			
25596900			13319 11256 12702 10624	23419 12702 10624 17639	17701 10624 17639 10175		
25745042						19418 27438 13090 10700	16716 13090 10700 27907
25778706				2652 12845 18592 5173	4077 18592 5173 7538	6374 5173 7538 8156	
25833774						6378 28741 31088 10881	947 31088 10881 11938
25916203	-403 3469 4477 7233						
25945084					10419 10627 11420 8948	10218 11420 8948 11708	
25962612			-3549 5316 10849 5662				
25995421							2840 9353 7917 8755
26026155				5118 5635 5665 5688	2871 5665 5688 3607		
44018711					10953 8643 10089 12513	9945 10089 12513 7034	6629 12513 7034 11095
45352861		-743 2347 2325 1707					
46349324				11852 5906 6185 18003	18071 6185 18003 10277	18984 18003 10277 13576	6314 10277 13576 8320
46351906				9039 9690 4902 13626	15712 4902 13626 5960		

Podnik	2000	2001	2002	2003	2004	2005	2006
46353836	766 -1630 611 10339	7781 611 10339 8104	15351 10339 8104 11106	13869 8104 11106 11808	20308 11106 11808 15801	17413 11808 15801 8284	12456 15801 8284 8145
46356819	-1763 4084 -35 5501	-1663 -35 5501 4119					
46357254				12087 32344 25698 40339	20839 40339 40339 33551	25594 40339 33551 24994	23107 33551 24994 26793
46357394				4601 2701 2938 6750	4945 2938 6750 4067	4601 6750 4067 3061	1651 4067 3061 3851
46678450					-4032 11980 9426 7056	-6187 9426 7056 13064	-7645 7056 13064 14165
46884335						8797 11006 14972 4170	
46969811			3771 7344 6140 4239	6975 6140 4239 6380	3772 4239 6380 3186		
46983198				7984 5600 5913 10539	14878 5913 10539 17250	11100 10539 17250 6969	6983 17250 6969 10504
47048174	-6750 1514 -48 8890						
47048336				5650 9953 10917 5403	4708 10917 5403 11478		
47048603				16348 16315 14101 18755	15017 14101 18755 21488	11174 18755 21488 9608	8975 21488 9608 9599
47150564			3479 3945 2337 2430	4585 2337 2430 4080	3854 2430 4080 2858	4224 4080 2858 3052	
47452471						14232 18750 16359 13693	12584 16359 13693 20793
47468297						22412 11741 25819 10863	23292 25819 10863 12684
47468629						-651 13749 3457 6455	-5792 3457 6455 7755
47470186					8814 4484 9616 6915		

Podnik	2000	2001	2002	2003	2004	2005	2006
47666102						2864 2463 3764 2915	2519 3764 2915 3441
47666145				-2457 1225 2155 8729	-1326 2155 8729 5860	-7170 8729 5860 -2811	
47673231	-4282 -631 1102 -3265	-1537 1102 -3265 2463	-1480 -3265 2463 2264	-90 2463 2264 2395	-276 2264 2395 2378	-424 2395 2378 1475	
47673656			16605 8383 13144 8622	20809 13144 8622 16863	26270 8622 16863 20018	18054 16863 20018 5351	16322 20018 5351 14030
47719532		7422 7916 10115 15244	14408 10115 15244 17208	23811 15244 17208 20322	24871 17208 20322 18193	20529 20322 18193 11453	14568 18193 11453 10692
47904933						-1968 5002 2992 -2113	
47906961				9511 11114 8081 7259	5741 8081 7259 8459	3282 7259 8459 5638	1816 8459 5638 5828
47912961	13356 33673 28055 -3780	5061 28055 -3780 26541	2404 -3780 26541 26556		23500 26556 29786 28655		20182 28655 26902 37054
48172936			4837 3780 2927 3306	3786 2927 3306 4652	2100 3306 4652 3173	1794 4652 3173 2971	
48173169			3948 3971 3327 3257	3358 3327 3257 3704	3692 3257 3704 4509		
48173291				22375 33051 28007 24649	30746 28007 24649 35186	29852 24649 35186 19810	24081 35186 19810 18703
48908754			6004 7270 6575 5744	4942 6575 5744 6813	4213 5744 6813 4873	2936 6813 4873 4815	1015 4873 4815 5754
49018345	10341 3696 6166 8183	11166 6166 8183 11746	12617 8183 11746 8091	9125 11746 8091 8565	21436 8091 8565 26990	21931 8565 26990 6364	31814 26990 6364 12687
49023314				3382 3527 2246 2592	2154 2246 2592 1553	2257 2592 1553 2876	
49050575				15486 11926 9010 9010	9284 9010 9010 5353	5744 9010 5353 5330	
49454072				-104 3446 304 2597	-2649 304 2597 881	-1339 2597 881 2212	-2063 881 2212 1762

Podnik	2000	2001	2002	2003	2004	2005	2006
49788183							7202 9554 5175 10394
49788744		-2262 12768 21905 23643					
49791265	139 5 -883 1737	974 -883 1737 914					
49812114							809 3518 1868 3078
49814478	3144 2564 354 4712	4429 354 4712 4033	5632 4712 4033 3863	5639 4033 3863 4650	3374 3863 4650 3977	314 4650 3977 2410	
49970763						39579 27958 37563 21343	27454 37563 21343 15837
49971492	-19374 -1646 -95 16849	-4655 -95 16849 14068	-681 16849 14068 7512				
49975552				8937 14957 14498 15378			
60067918	776 -481 -938 4886	4428 -938 4886 3783	6194 4886 3783 1975	4427 3783 1975 4611	6706 1975 4611 6525	7408 4611 6525 3208	7059 6525 3208 3229
60112336				1107 7105 5881 6713	-2353 5881 6713 4972	-253 6713 4972 9246	4310 4972 9246 12809
60825677	5306 4576 2735 8400						19631 27182 19829 29487
60838451			2653 3708 4945 3191	3765 4945 3191 3919	5819 3191 3919 7381	5042 3919 7381 3913	5446 7381 3913 5833
61974986				2048 22447 9831 31405	-1360 9831 31405 33307	3734 31405 33307 18764	
62619527					19619 14109 13065 10609	13915 13065 10609 5516	
63470381					-1960 20038 20042 13142		
63474964							1775 6458 4124 8087

Podnik	2000	2001	2002	2003	2004	2005	2006
63493021				10105 8184 11368 12700			
63495392						7138 16910 16155 4985	
64356370	1121 11393 8676 12459	-2753 8676 12459 11677					
64506576	-5436 7828 4711 8883	-15931 4711 8883 4021	-9925 8883 4021 5487	-4078 4021 5487 8602	-5839 5487 8602 -9608	-15514 8602 -9608 -2254	-16534 -9608 -2254 5016
64506843				14112 21131 21716 19057	8186 21716 19057 19115		
64610047					-5920 7810 9108 3521	639 9108 3521 18579	
64789462				25589 17714 4055 30914	31759 4055 30914 24569		
64834646	-6055 12994 9861 31585						
65006038	-6694 -3278 -1542 -1788	-3771 -1542 -1788 2980	783 -1788 2980 3881				
65278844			19326 26226 25610 22512	43912 25610 22512 46123			

Pramen: vlastní zpracování

## Příloha B

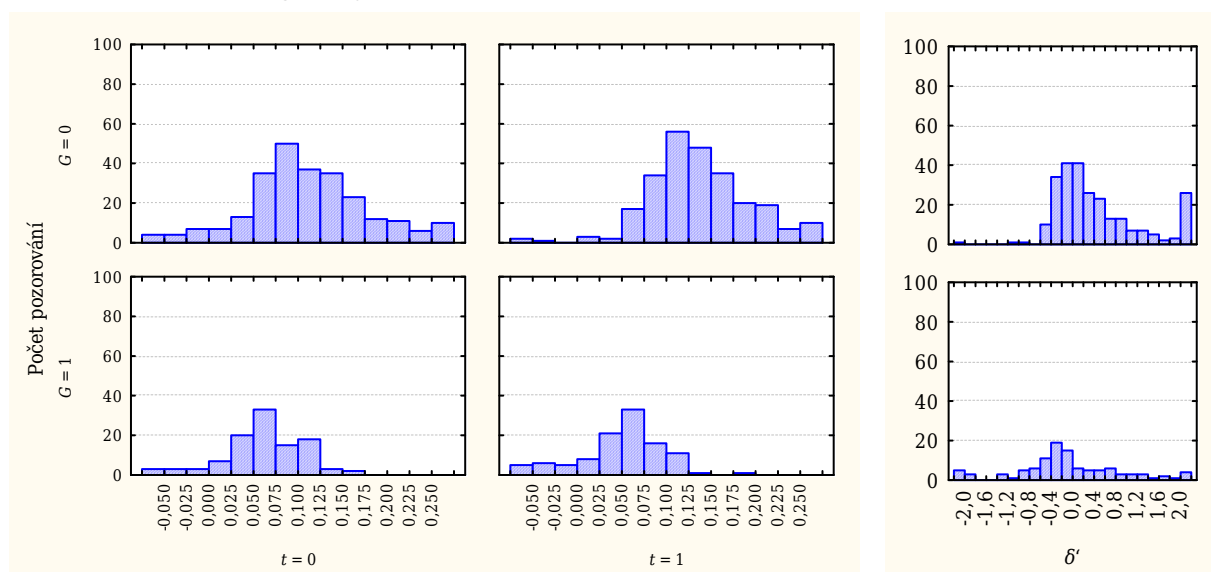
Příloha obsahuje základní statistické charakteristiky ukazatelů, které sloužily jako vstupy pro vícerozměrné klasifikační metody. Jejich hodnoty jsou uvedeny jak pro období  $t = 0$  a  $t = 1$ , tak i pro modifikovanou relativní změnu vždy zvlášť pro skupinu podniků ohrožených finanční tísni a pro skupinu prosperujících podniků. Hodnoty lze využít například při normalizaci nových pozorování při aplikaci modelu neuronové sítě. Z grafů intervalových četností si lze vytvořit představu o separabilitě skupin (opět pro obě období i změnu mezi obdobími).

Tab. B-1 Základní statistické charakteristiky ukazatele  $CF/A$

Období	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
Průměr	0,0604	0,1121	0,0474	0,1383	-0,1622	1,1079
Sm. odch.	0,0491	0,0782	0,0511	0,0612	1,9031	3,6237
Var. koef.	81,230	69,781	107,61	44,257	-1173,4	327,09
Šikmost	-1,1482	0,8354	-0,9976	0,5469	-2,7013	5,0690
Špičatost	3,3365	4,7178	1,9976	2,5561	25,7217	28,138
Minimum	-0,1508	-0,1791	-0,1508	-0,0781	-13,5405	-2,1832
Dolní kv.	0,0366	0,0710	0,0284	0,1005	-0,4939	-0,1021
Medián	0,0621	0,1043	0,0560	0,1284	-0,1989	0,1829
Horní kv.	0,0901	0,1496	0,0780	0,1657	0,5084	0,7856
Maximum	0,1601	0,5488	0,1770	0,3959	8,0091	27,469

Pramen: vlastní zpracování

Obrázek B-1 Histogramy ukazatele  $CF/A$



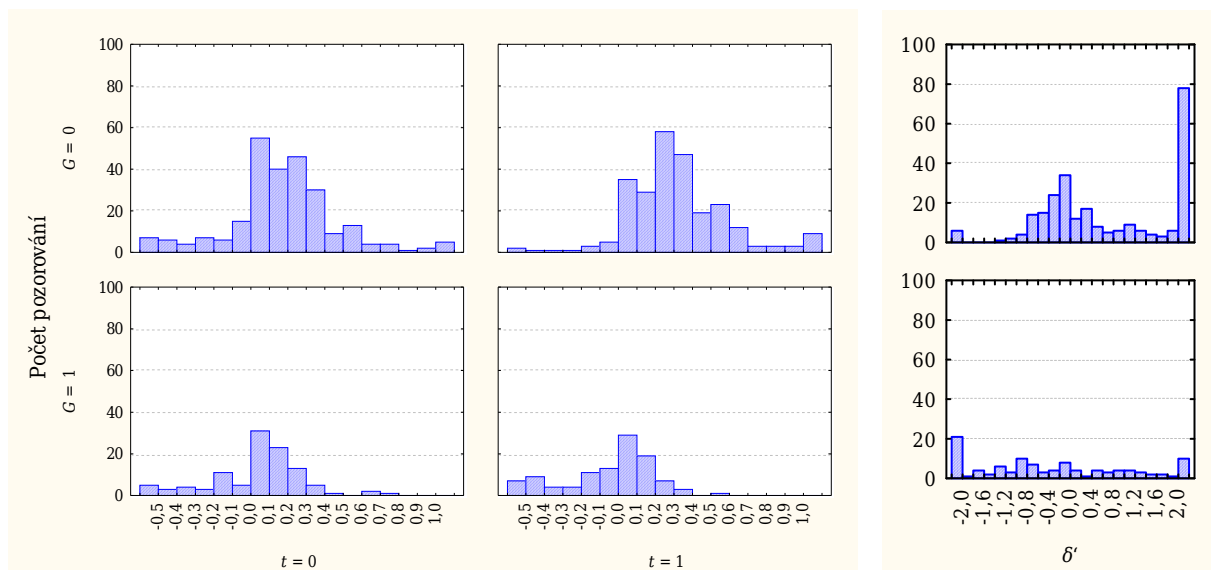
Pramen: vlastní zpracování

Tab. B-2 Základní statistické charakteristiky ukazatele  $ZD/Mz$

Období	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
Průměr	0,0275	0,1687	-0,0645	0,3212	-1,1020	3,3811
Sm. odch.	0,2676	0,3268	0,2639	0,2865	5,3817	21,931
Var. koef.	974,24	193,69	-409,12	89,194	-488,33	648,65
Šikmost	-0,8318	-0,1737	-0,9823	0,9620	-1,4212	14,309
Špičatost	2,4963	4,0327	1,0777	3,6221	13,706	218,15
Minimum	-0,9475	-1,4511	-0,9475	-0,6044	-29,854	-20,755
Dolní kv.	-0,0892	0,0310	-0,1618	0,1493	-1,5517	-0,2368
Medián	0,0535	0,1670	0,0097	0,2771	-0,6210	0,3646
Horní kv.	0,1800	0,3103	0,1149	0,4448	0,7101	2,4628
Maximum	0,7781	1,3430	0,5289	1,6558	23,801	338,62

Pramen: vlastní zpracování

Obrázek B-2 Histogramy ukazatele  $ZD/Mz$



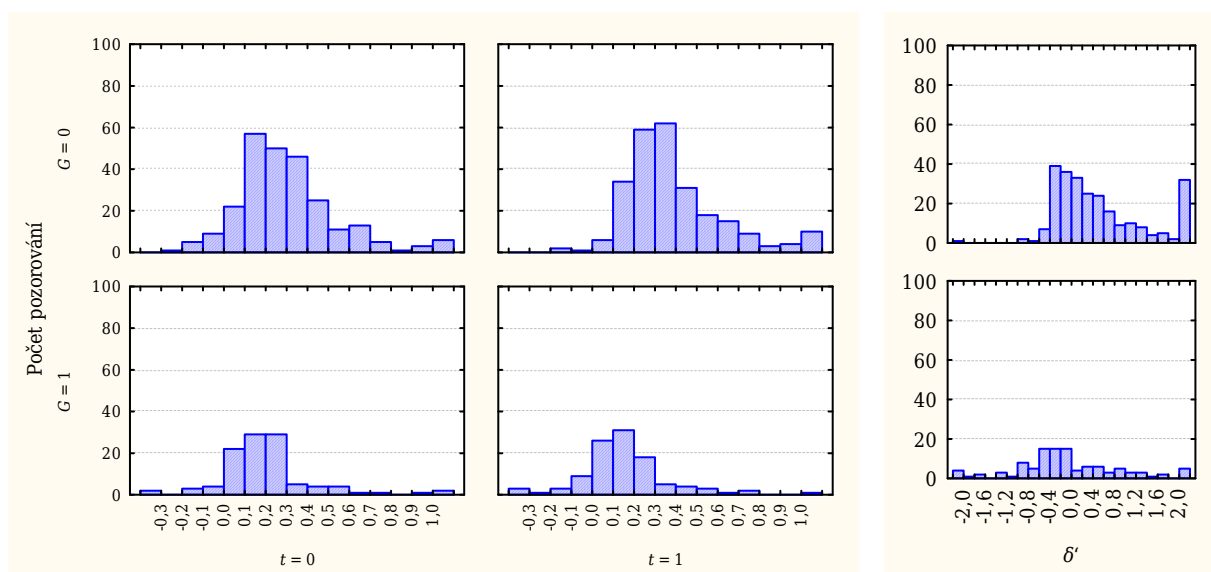
Pramen: vlastní zpracování

Tab. B-3 Základní statistické charakteristiky ukazatele  $CF/CZ$

Období	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
Průměr	0,2049	0,3096	0,1585	0,3988	-0,1418	1,4385
Sm. odch.	0,2429	0,2802	0,2161	0,2839	2,0253	5,7435
Var. koef.	118,54	90,508	136,34	71,186	-1427,9	399,27
Šikmost	2,4553	2,5191	1,1069	2,4703	-1,9667	7,7922
Špičatost	12,020	12,187	4,3342	10,114	17,871	71,005
Minimum	-0,3313	-0,2055	-0,3961	-0,1990	-11,350	-2,8421
Dolní kv.	0,0867	0,1483	0,0440	0,2295	-0,5268	-0,1365
Medián	0,1800	0,2691	0,1436	0,3309	-0,2103	0,2606
Horní kv.	0,2501	0,4026	0,2394	0,4614	0,5160	0,9407
Maximum	1,6360	2,1425	1,1467	2,1425	8,8430	62,727

Pramen: vlastní zpracování

Obrázek B-3 Histogramy ukazatele  $CF/CZ$



Pramen: vlastní zpracování

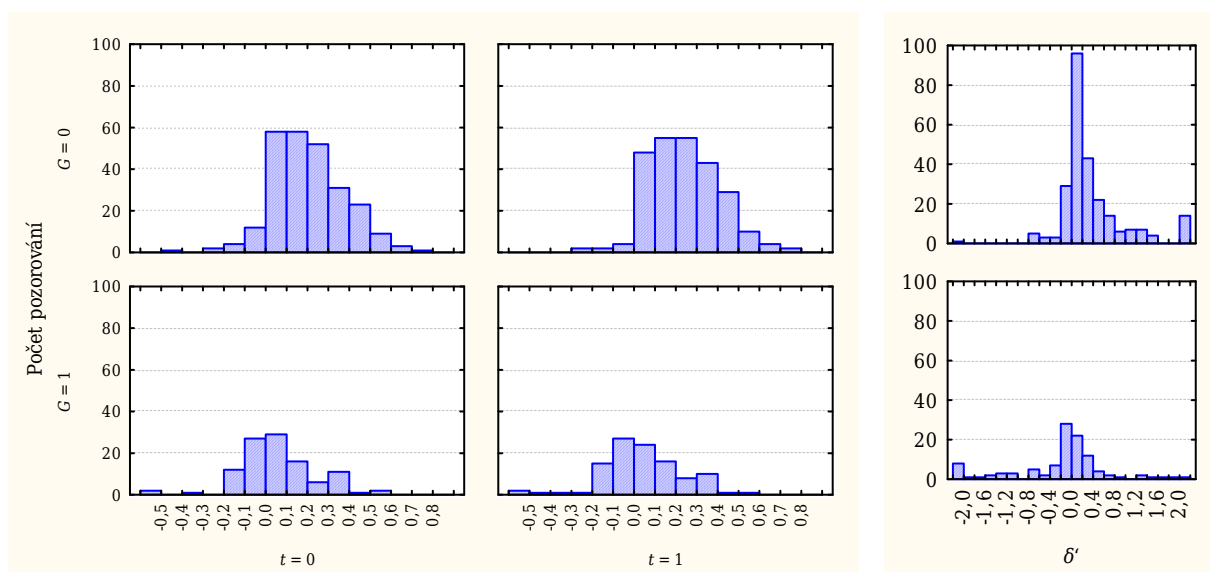


Tab. B-4 Základní statistické charakteristiky ukazatele  $DR$

Období	$t = 0$		$t = 1$		$\delta'$	
Skupina	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
Průměr	0,0508	0,2007	0,0360	0,2391	-0,3678	0,7827
Sm. odch.	0,1814	0,1738	0,1862	0,1677	1,6988	3,4168
Var. koef.	357,03	86,610	517,53	70,158	-461,89	436,54
Šikmost	0,1137	0,1254	-0,1464	0,2894	-5,4534	8,1956
Špičatost	1,2899	0,8495	1,2255	0,4394	41,891	75,897
Minimum	-0,5084	-0,4933	-0,5209	-0,2766	-14,2193	-5,0562
Dolní kv.	-0,0590	0,0769	-0,0686	0,1145	-0,3848	0,0748
Medián	0,0293	0,1851	0,0177	0,2410	-0,0325	0,1876
Horní kv.	0,1607	0,3063	0,1369	0,3354	0,1893	0,4450
Maximum	0,5625	0,7503	0,5729	0,7936	2,2454	38,038

Pramen: vlastní zpracování

Obrázek B-4 Histogramy ukazatele  $DR$



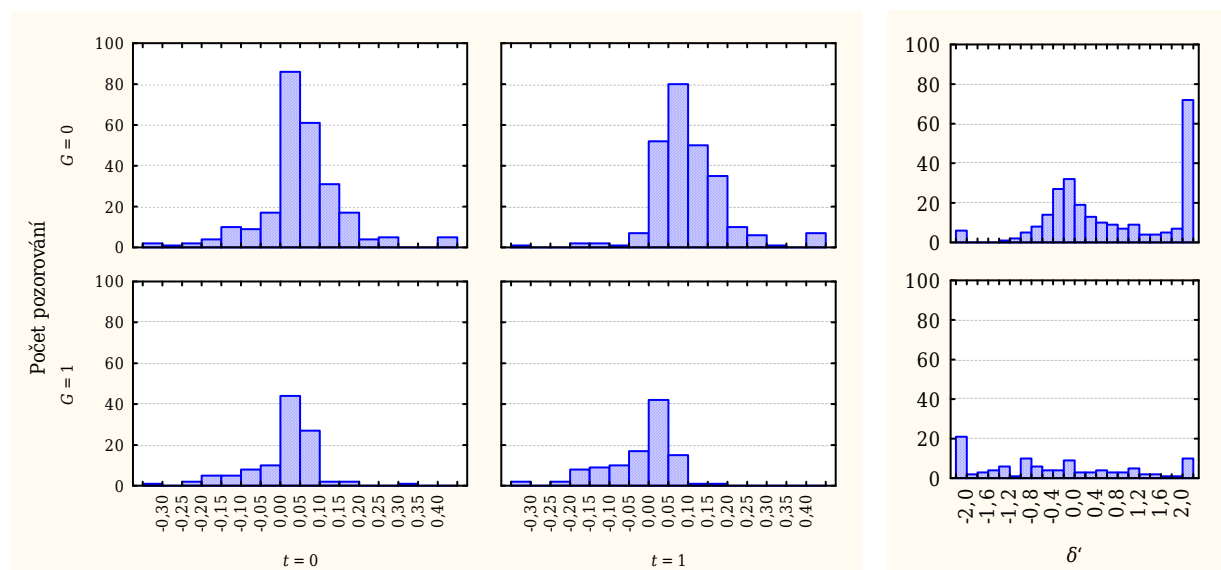
Pramen: vlastní zpracování

Tab. B-5 Základní statistické charakteristiky ukazatele ZD/W

Období	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
Průměr	0,0053	0,0553	-0,0239	0,1072	-1,3056	3,4357
Sm. odch.	0,0902	0,1202	0,0902	0,1320	5,6299	23,412
Var. koef.	1686,6	217,54	-376,52	123,13	-431,21	681,43
Šikmost	-0,9231	1,2690	-1,2199	4,5585	-2,2171	13,790
Špičatost	3,6244	9,1759	1,9455	39,193	14,322	204,87
Minimum	-0,3768	-0,3730	-0,3768	-0,3185	-31,797	-26,818
Dolní kv.	-0,0341	0,0121	-0,0624	0,0486	-1,5957	-0,1917
Medián	0,0162	0,0484	0,0040	0,0865	-0,5888	0,4002
Horní kv.	0,0556	0,0979	0,0367	0,1423	0,6238	2,3258
Maximum	0,3078	0,8116	0,1610	1,3932	21,394	355,44

Pramen: vlastní zpracování

Obrázek B-5 Histogramy ukazatele ZD/W



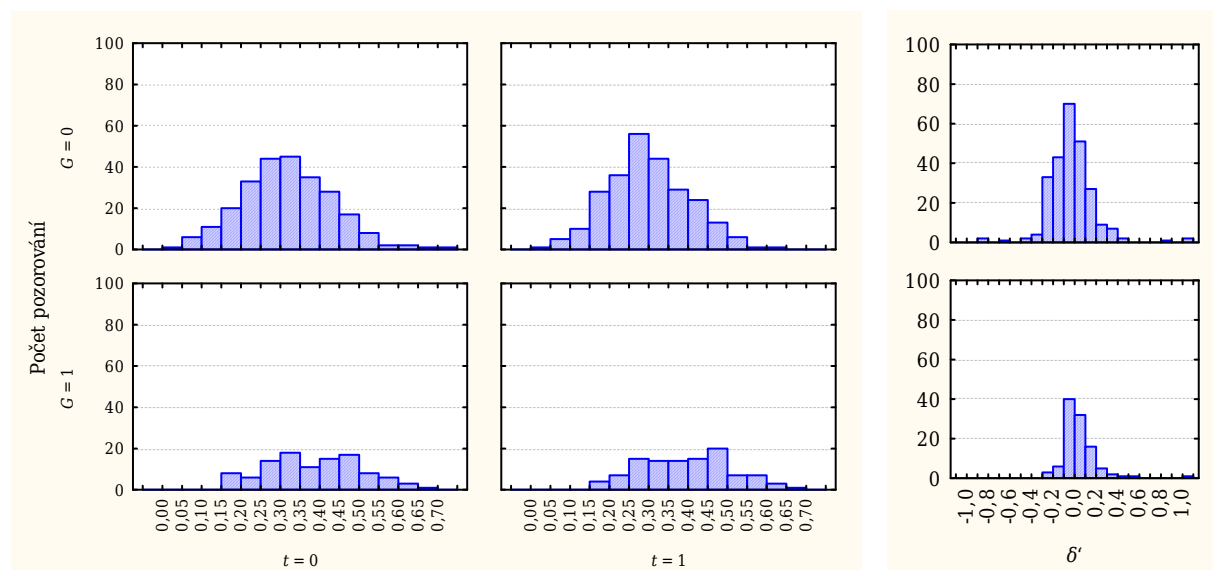
Pramen: vlastní zpracování

Tab. B-6 Základní statistické charakteristiky ukazatele

Období	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
Průměr	0,3854	0,3211	0,3946	0,2991	0,0440	0,0085
Sm. odch.	0,1163	0,1464	0,1111	0,1039	0,1975	0,7372
Var. koef.	30,163	45,592	28,159	34,732	448,77	8710,1
Šikmost	0,0641	4,0445	0,0294	0,1256	4,6465	14,202
Špičatost	-0,6356	39,4645	-0,6217	-0,0965	33,444	217,42
Minimum	0,1569	0,0132	0,1569	0,0132	-0,2787	-0,8865
Dolní kv.	0,2992	0,2444	0,3012	0,2310	-0,0585	-0,1482
Medián	0,3880	0,3108	0,3954	0,2938	0,0095	-0,0438
Horní kv.	0,4734	0,3970	0,4762	0,3732	0,0990	0,0685
Maximum	0,6563	1,7936	0,6563	0,6062	1,5680	11,270

Pramen: vlastní zpracování

Obrázek B-6 Histogramy ukazatele  $Zas/V$



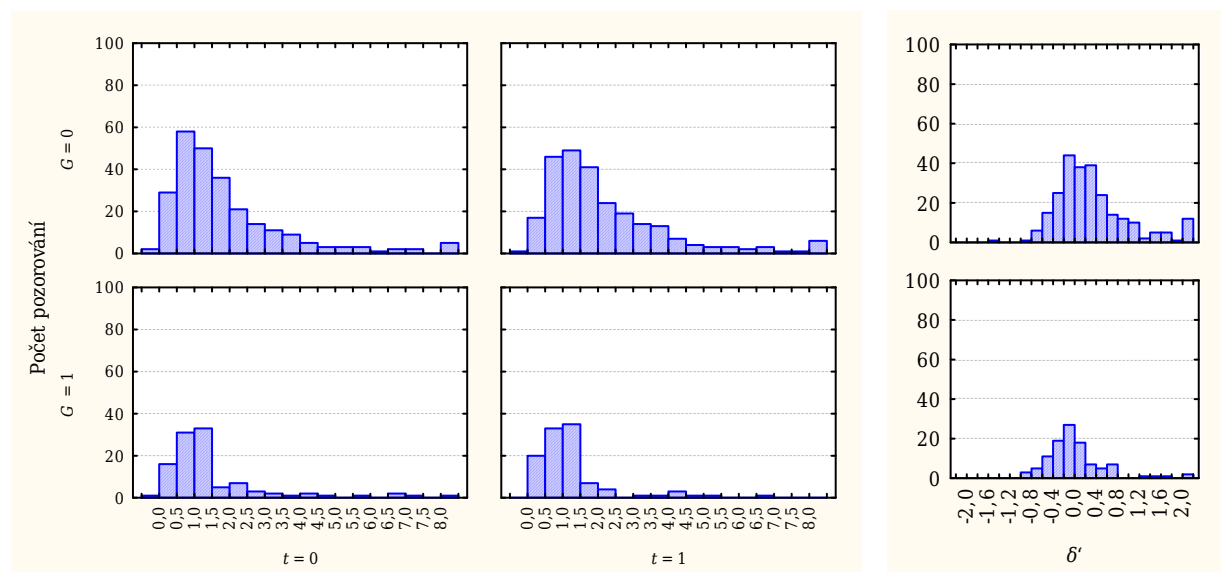
Pramen: vlastní zpracování

Tab. B-7 Základní statistické charakteristiky ukazatele  $L2$

Období	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
Průměr	1,4980	1,7862	1,2198	2,2387	0,0810	0,5356
Sm. odch.	1,5179	2,7971	1,1053	2,1529	0,9922	1,7434
Var. koef.	101,33	156,59	90,609	96,166	1224,5	325,52
Šikmost	2,6090	-5,6618	2,7592	3,3906	5,1441	5,4203
Špičatost	7,2520	79,130	8,8840	18,322	31,368	35,265
Minimum	-0,0632	-31,263	0,1047	-2,2756	-0,9343	-1,4306
Dolní kv.	0,6643	0,8326	0,5944	0,9976	-0,3116	-0,1103
Medián	1,0615	1,3336	1,0244	1,7060	-0,0894	0,1875
Horní kv.	1,4789	2,3006	1,3618	2,6906	0,1854	0,5836
Maximum	8,2380	14,884	6,8686	18,571	6,6220	15,907

Pramen: vlastní zpracování

Obrázek B-7 Histogramy ukazatele  $L2$



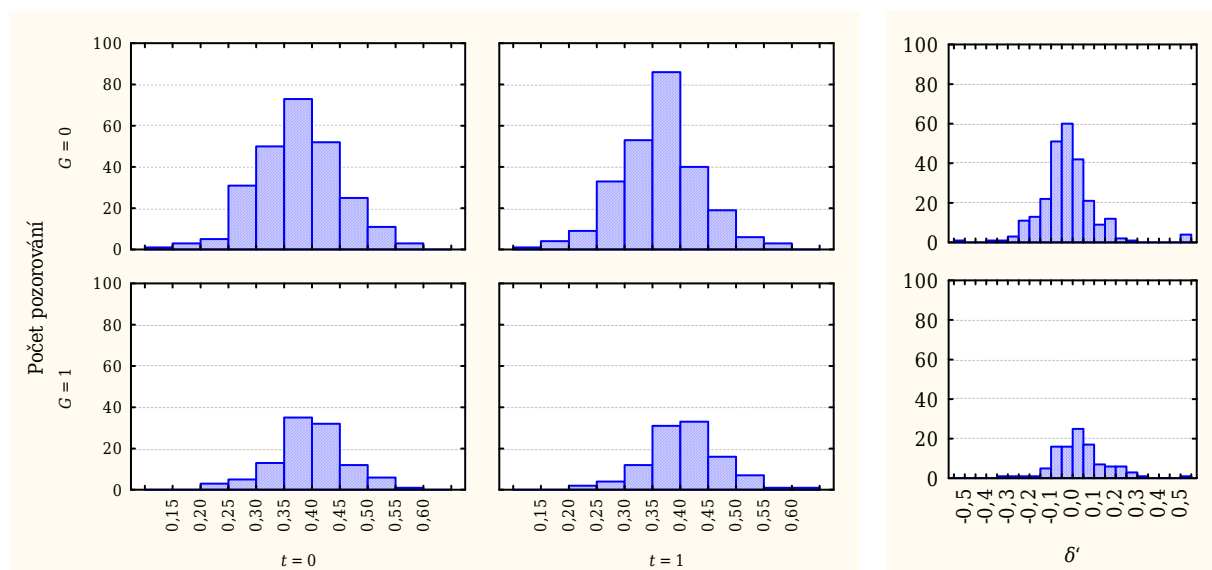
Pramen: vlastní zpracování

Tab. B-8 Základní statistické charakteristiky ukazatele  $Mt/V$

Období	$t = 0$		$t = 1$		$\delta'$	
	$G = 1$	$G = 0$	$G = 1$	$G = 0$	$G = 1$	$G = 0$
Průměr	0,3979	0,3746	0,4091	0,3628	0,0377	-0,0211
Sm. odch.	0,0684	0,0755	0,0688	0,0725	0,1416	0,1533
Var. koef.	17,180	20,162	16,827	19,976	375,28	-727,23
Šikmost	-0,1107	-0,0642	0,0623	-0,0245	1,9688	3,3097
Špičatost	0,5045	0,3605	0,7259	0,5474	11,176	25,862
Minimum	0,2161	0,1492	0,2161	0,1492	-0,3093	-0,5296
Dolní kv.	0,3638	0,3294	0,3665	0,3172	-0,0424	-0,0879
Medián	0,3946	0,3754	0,4122	0,3652	0,0165	-0,0308
Horní kv.	0,4416	0,4202	0,4464	0,4022	0,0908	0,0349
Maximum	0,5734	0,5793	0,6087	0,5793	0,8825	1,3216

Pramen: vlastní zpracování

Obrázek B-8 Histogramy ukazatele  $Mt/V$



Pramen: vlastní zpracování