

MULTIFAKTOROVÁ ANALÝZA DOPRAVNÍ NEHODOVOSTI

metodika provádění



Tato metodika byla zpracována v rámci výzkumného projektu „Identifikace a řešení kritických míst a úseků v síti pozemních komunikací, které svým uspořádáním stimulují nezákonné a nepřiměřené chování účastníků silničního provozu“ (IDEKO) Programu bezpečnostního výzkumu ČR na léta 2010-2015 Ministerstva vnitra (VG20112015013).

Oponenti byli Doc. Ing. Josef Kocourek, Ph.D. (České vysoké učení technické v Praze, Fakulta dopravní) a Ing. Vlasta Michková (Ředitelství silnic a dálnic ČR).

Metodika byla certifikována osvědčením Ministerstva dopravy č.j. 113/2014-710-VV/1 ze dne 5. 12. 2014.

ODPOVĚDNÝ ŘEŠITEL

Ing. Radim Striegler

AUTORSKÝ TÝM

Ing. Jiří Ambros

Ing. Veronika Valentová

Ing. Radim Striegler

Ing. Pavel Havránek

Ing. Eva Simonová

Ing. Martin Lipl

Ing. Lucie Vyskočilová

Ing. Petr Pokorný

Ing. Jan Novák

Mgr. Martin Kyselý

Mgr. Zbyněk Janoška

Ing. Ondřej Gogolín

OBSAH

1	Úvod	4
1.1	<i>Dopravní nehodovost</i>	4
1.2	<i>Multifaktorová analýza</i>	4
1.3	<i>Vymezení rozsahu metodiky</i>	5
2	Teoretická část	7
2.1	<i>Konceptuální rámec</i>	7
2.2	<i>Explorační analýza</i>	8
2.3	<i>Analýza závislostí</i>	9
2.3.1	<i>Korelace</i>	10
2.3.2	<i>Lineární regrese</i>	12
2.3.3	<i>Logistická regrese</i>	12
2.3.4	<i>Zobecněný lineární model</i>	13
2.4	<i>Aplikace v kontextu nehodovosti</i>	14
3	Praktická část	16
	<i>Příklad A – Křižovatky</i>	16
A.1	<i>Příprava dat</i>	16
A.2	<i>Explorační analýza</i>	16
A.3	<i>Analýza závislostí</i>	19
A.3.1	<i>Korelace</i>	19
A.3.2	<i>Logistická regrese</i>	19
A.3.3	<i>Zobecněný lineární model</i>	20
	<i>Příklad B – Mezikřižovatkové úseky</i>	21
B.1	<i>Příprava dat</i>	21
B.2	<i>Explorační analýza</i>	22
B.3	<i>Analýza závislostí</i>	26
B.3.1	<i>Korelace</i>	26
B.3.2	<i>Logistická regrese</i>	26
B.3.3	<i>Zobecněný lineární model</i>	27
	<i>Závěr</i>	29
4	Přínosy metodiky	31
4.1	<i>Novost postupu</i>	31
4.2	<i>Popis uplatnění certifikované metodiky</i>	32
4.3	<i>Ekonomické aspekty</i>	33
5	Reference	34
5.1	<i>Seznam použité související literatury</i>	34
5.2	<i>Seznam publikací, které předcházely metodice</i>	35
5.3	<i>Výstupy z originální práce</i>	35

1 Úvod

Tato certifikovaná metodika se zabývá multifaktorovou analýzou dopravní nehodovosti – jejím cílem je poskytnout teoretický i praktický podklad pro její provádění. Multifaktorovou analýzu lze provádět např. na úrovni krajské silniční sítě, s cílem identifikace kritických faktorů, na které je vhodné zaměřit dopravně bezpečnostní činnost. Metodika je určena pro ty, kdo budou tyto analýzy zadávat, provádět a/nebo interpretovat, tedy:

- zodpovědné pracovníky správce silniční sítě (ŘSD, SÚS, městských odborů dopravy),
- zodpovědné pracovníky Ministerstva dopravy nebo Ministerstva vnitra,
- vědecké pracovníky výzkumných institucí nebo vysokých škol.

Metodika vznikla v rámci výzkumného projektu IDEKO („Identifikace a řešení kritických míst a úseků v síti pozemních komunikací, které svým uspořádáním stimulují nezákonné a nepřiměřené chování účastníků silničního provozu“), řešeného Centrem dopravního výzkumu, v.v.i. (CDV). Skladba autorského týmu byla následující:

- statistická část: Ing. Jiří Ambros, Ing. Veronika Valentová, Ing. Jindřich Frič, Ph.D.
- dopravně-inženýrská část: Ing. Radim Striegler, Ing. Pavel Havránek, Ing. Eva Simonová, Ing. Martin Lipl, Ing. Lucie Vyskočilová, Ing. Petr Pokorný, Ing. Jan Novák
- příprava podkladů: Mgr. Martin Kyselý, Mgr. Zbyněk Janoška, Ing. Ondřej Gogolín

Autoři děkují Ing. Michalu Kadlecovi (ACREA CR, spol. s.r.o.) za konzultace a připomínky k praktické části.

Cílem multifaktorové analýzy dopravní nehodovosti je identifikace kritických vlivů a kvantifikace míry jejich působení na vznik dopravních nehod. Metodika obsahuje dvě hlavní části: v první stručné části (teoretické) je představen princip vybraných metod; důraz a rozsah je kladen především na druhou (praktickou) část, kde jsou metody předvedeny na dvou příkladech.

1.1 Dopravní nehodovost

Dopravní nehodovost je ovlivňována celou řadou faktorů, které se nejčastěji dělí do tří skupin:

- účastníci provozu (např. pozornost nebo používání bezpečnostních pásů)
- dopravní prostředky (např. technický stav vozidla)
- dopravní infrastruktura a její okolí (např. uspořádání komunikace)

(zkráceně *řidič, vozidlo, prostředí*)

V historii byly tyto faktory často zkoumány ve smyslu *příčin* – bylo zjišťováno, zda jsou nehody častěji způsobené chybou řidiče, vozidla nebo prostředí. Ve skutečnosti však nehoda většinou nebývá důsledkem jediné příčiny – spíše se jedná o souhru *více vlivů (faktorů)*.

1.2 Multifaktorová analýza

Pojem *multifaktorová analýza* označuje soubor postupů určených k analýze jevů, na které působí více než 1 vliv (faktor) – tyto jevy se proto nazývají multifaktorové, víceproměnné nebo vícerozměrné. Mezi takové jevy patří i dopravní nehodovost.

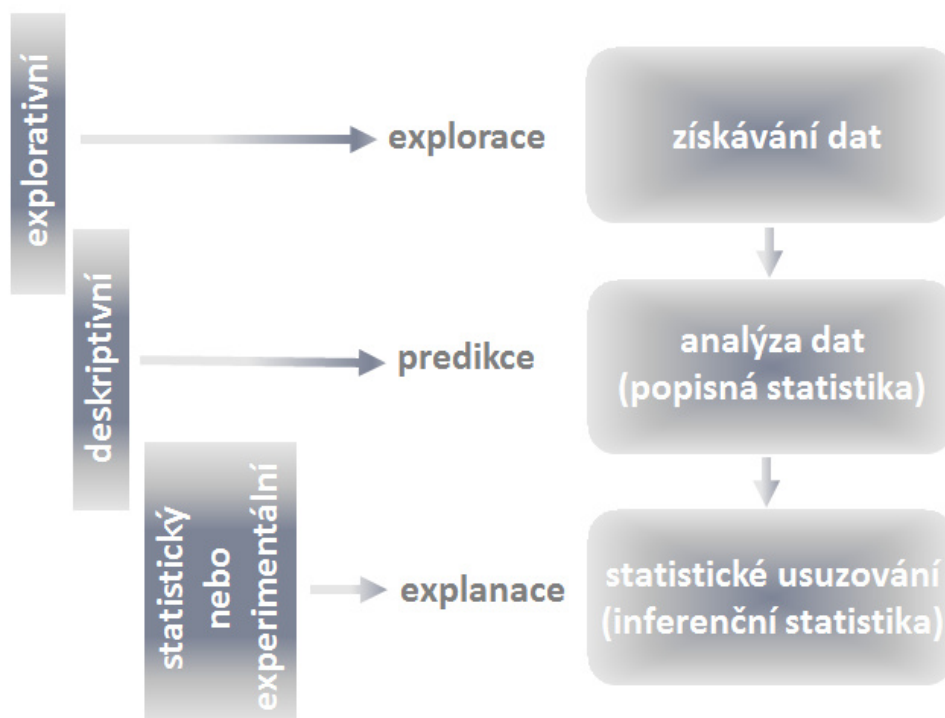
Proces analýzy lze rozdělit do následujících kroků: [14]

- 1) Výběr analyzovaného jevu (např. roční četnost nehod na křižovatce, nehody za mokra, nehody způsobené nedáním přednosti v jízdě).
- 2) Volba vlivů, o kterých se předpokládá, že nehodovost ovlivňují a zjištění jejich hodnot (vlastním průzkumem nebo ze sekundárních dat).
- 3) Provedení analýzy, které ověří, zda zvolené veličiny nehodovost opravdu ovlivňují a pokud ano, tak jakou měrou.

Multifaktorová analýza se používá na plošné hodnocení – např. na celostátní, krajské nebo městské silniční síti. Jednotkou analýzy jsou např. křižovatky nebo mezikřižovatkové úseky. Vstupem jsou proto data, která popisují všechny jednotky ve vybrané oblasti (doporučuje se minimální počet 20): nejčastěji se využívají nehodová data Policie ČR, dále údaje o dopravním zatížení z Celostátního sčítání dopravy a údaje o silniční síti ze Silniční databanky ŘSD ČR; doplňkově lze provést vlastní průzkum (např. rychlosti, dopravní značení). Výstupem analýzy je rovnice, která obsahuje jednotlivé statisticky významné proměnné (tj. rizikové faktory) a koeficienty udávající míru jejich vlivu na nehodovost či na zkoumanou příčinu nehod na dané síti.

1.3 Vymezení rozsahu metodiky

Multifaktorová analýza využívá řadu statistických postupů (technik), předložená metodika však nutně pokrývá pouze jejich výběr. Pro vysvětlení lze uvést následující schéma (Obr. 1).



Obr. 1 Přehled výzkumných plánů, cílů a částí praxe statistiky [7, 10, 11]

Schéma představuje základní výzkumné plány (úrovně), jejich záběry se částečně překrývají: [11]

- *explorativní* – pouhý popis zkoumaného jevu, vhodný pro formulaci hypotéz
- *deskriptivní* – popis zkoumaného jevu včetně kvantifikace (statistik výskytu)
- *statistický nebo experimentální* – popis vztahů mezi jevy, jedná se o souvislost (korelační vztahy – u statistického přístupu) nebo příčinnost (kauzální vztahy – u experimentu)

Výzkumné plány vedou k hlavním cílům vědeckého poznání: [7]

- *explorace* – popis a utřídění problému (deskripce a klasifikace)
- *predikce* – předpovědění vztahů mezi jevy
- *explanace* – vysvětlení těchto vztahů

Těmto konceptům odpovídají příslušné části praxe statistiky: [10]

- *získávání dat*
- *analýza dat* – souhrny, grafy a další matematické nástroje popisné statistiky
- *statistické usuzování* – testování hypotéz včetně intervalů spolehlivosti a dalších pravděpodobnostních postupů inferenční statistiky

Vědecky nejhodnotnější závěry o příčinném působení (což je ideálním cílem multifaktorové analýzy) lze získat prostřednictvím experimentálního výzkumného plánu – ten však je v oboru dopravy prakticky nerealizovatelný. Další možností je tedy statistický výzkumný plán, jenž s pomocí statistických nástrojů vede k vysvětlení korelačních vztahů (souvislostí) zkoumaných jevů.

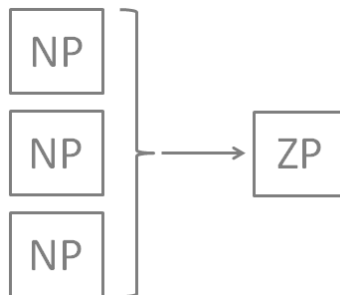
Platí však, že uvedené výzkumné plány a jejich nástroje a cíle se částečně překrývají. Dále platí, že se často budují postupně: prvním krokem je vždy sběr dat a jejich základní explorační (průzkumová) analýza. Závěrem je, že předložená metodika se soustřeďuje především na statistický výzkumný plán (explanaci); okrajově zmiňuje i explorační a predikci. Jedná se však spíše o výtah (nebo „rozcestník“) v uvedených oblastech statistiky, nejedná se o podrobný učební text ani manuál ke statistickému software. Na podrobnosti jsou v textu odkazy, dále metodika obsahuje praktickou část s okomentovanými ukázkovými příklady, ve kterých byl použit jeden z rozšířených statistických programů.

2 Teoretická část

Teoretická část v úvodu představuje konceptuální rámec multifaktorových analýz. Dále jsou představeny vybrané techniky explorační analýzy a analýzy závislostí. Závěrem je uvedena jejich aplikace v kontextu nehodovosti.

2.1 Konceptuální rámec

Cílem metodiky je identifikace kritických vlivů a zjištění míry jejich působení na vznik dopravních nehod či jejich určité příčiny (nedání přednosti v jízdě, nesprávné předjíždění, ...) – tj. kvantifikace vlivu jednotlivých nehodových faktorů. Za tím účelem budou využity přístupy kvantitativního výzkumu. Ten se zabývá vztahy mezi proměnnými, jejichž hodnoty mohou být čísla nebo kategorie. *Závisle proměnná* bude nehodovost (např. ve formě roční četnosti nehod na křižovatce nebo roční hustotě nehod na 1 km délky mezikřižovatkového úseku, mohou být zkoumány také pouze nehody určité příčiny). *Nezávisle proměnné* budou jednotlivé nehodové faktory. Obr. 2 schematicky znázorňuje obecný konceptuální rámec tohoto typu výzkumu: závisle proměnná (ZP) je ovlivňována několika nezávislými proměnnými (NP).



Obr. 2 Obecný konceptuální rámec [14]

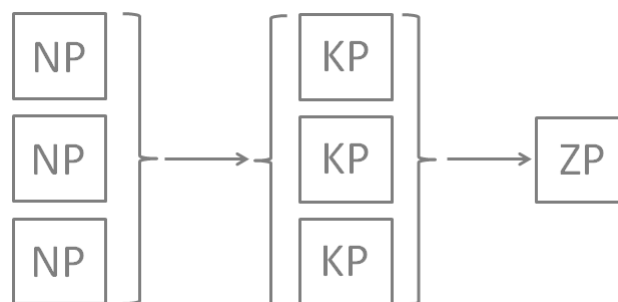
Teoreticky by nezávisle proměnné měly být *příčiny* a závisle proměnné by měly být jejich *důsledkem*. Toto však platí jen v případě čistě *experimentálního* výzkumu, kdy se pomocí manipulace s nezávislými proměnnými zjišťuje vliv na závisle proměnnou, přičemž tento vliv není zkreslen žádnou jinou proměnnou. Toho lze docílit v laboratorním výzkumu (v dopravním kontextu např. na řídicím simulátoru), obecně však v dopravě tento předpoklad neplatí. Je tedy vhodnější používat pojem *vlivy* nebo *faktory*.

Jak bylo uvedeno v odstavci 1.1, faktory ovlivňující nehodovost lze rozdělit do tří základních skupin. Kombinace faktorů reprezentuje stav, kdy dojde k dopravní nehodě či má vliv na závažnost jejích následků. Některé z nich uvádí následující výčet:

- **Řidič**
 - Psychický stav (frustrace, rozčilení, nesoustředěnost...)
 - Fyzický stav (únava, zdravotní stav, zrakové omezení...)
 - Chování ostatních účastníků provozu (dodržování pravidel provozu, agresivní chování...)
- **Vozidlo**
 - Parametry vozidla (výkon motoru, počet hnaných náprav, počet airbagů, ABS, ESP ...)
 - Technický stav (stáří pneumatik, stav karoserie, funkčnost brzdového systému, funkčnost asistenčních systémů, funkčnost zádržných systémů...)

- *Prostředí*
 - Stavebně technický stav komunikace (stupeň opotřebení povrchu, podélné nebo příčné vlny, výtlučky neodstranitelné běžnou údržbou, únosnost vozovky, krajnic, mostů a mostních objektů a vybavení pozemní komunikace součástmi a příslušenstvím)
 - Dopravně technický stav komunikace (příčné uspořádání, příčný a podélný sklon, šířka a druh vozovky, směrové a výškové oblouky, rozhledové poměry, nadmožská výška)
 - Povětrnostní podmínky (slunečno, déšť, mlha, sněžení, vítr, námraza, ledovka...)
 - Denní doba (den, noc)

Obecně platí, že vlivy (tj. nezávisle proměnné) by se měly volit na základě teorie (např. fyzikální) [16]. V oblasti bezpečnosti dopravy však bohužel ustálené teorie chybí [5]. Dále platí, že v dopravě nelze manipulovat s proměnnými, dopravní jevy lze proto zkoumat pouze neexperimentálně – pozorováním (observací), odtud pojem *observační* výzkum. Při tomto výzkumu se využívá přirozená variabilita nezávisle proměnných. Přitom však působí celá řada tzv. matoucích proměnných, jejichž vliv je nutno statisticky odstranit – tzv. kontrolovat a to tak, že se kontrolní proměnné zapojí do analýzy spolu s ostatními nezávisle proměnnými. Obr. 3 znázorňuje konceptuální rámec se zohledněním kontrolní proměnné (KP).



Obr. 3 Konceptuální rámec s kontrolními proměnnými [14]

Obecný postup multifaktorové analýzy byl uveden v odstavci 1.2. S využitím uvedeného konceptuálního rámce lze přepsat následovně:

- 1) Volba závislé proměnné.
- 2) Volba nezávisle proměnné/proměnných (včetně kontrolních proměnných).
- 3) Provedení analýzy, ve které jsou zjištěny hodnoty nezávisle proměnných; kvantifikace vztahů mezi závislou proměnnou a nezávisle proměnnými.

2.2 Explorační analýza

Účelem explorační (průzkumové) analýzy dat (angl. *exploratory data analysis*, EDA) je odhalit jejich zvláštnosti ještě před vlastním (do jisté míry rutinním) statistickým zpracováním [12]. Mezi její základní prvky patří vizualizace dat (grafická nebo tabelární) a analýza reziduálních (odlehých) hodnot [10].

Jednou z technik explorační analýzy je *metoda hlavních komponent* (angl. *principal component analysis*, PCA). Ta se používá k redukci počtu proměnných pomocí tzv. hlavních komponent, které popisují variabilitu všech proměnných a vztahy mezi nimi. Hlavní

komponenty vznikají jako lineární kombinace původních proměnných. Zkoumání těchto nových proměnných místo původních nám může umožnit datům snadněji porozumět.

Dále platí, že nově vytvořené proměnné jsou vzájemně nekorelované – každá z nich tedy měří jinou vlastnost (dimenzi dat). Nové proměnné jsou uspořádány podle svého rozptylu, ty s nejnižším rozptylem lze proto vyloučit a dosáhnout tak úspornějšího popisu chování původních proměnných pomocí menšího počtu nových proměnných.

Velikosti rozptylů jsou určeny velikostí vzájemných korelací proměnných. Pokud všechny proměnné korelují absolutně, je možné celkový rozptyl zachytit pouze jednou hlavní komponentou. Jestliže jsou naopak všechny korelace nízké, k vysvětlení celkového rozptylu je potřeba tolik komponent, kolik je původních proměnných.

Postup analýzy hlavních komponent je následující: [10]

- 1) Počáteční analýza za účelem zjištění existence vztahu mezi proměnnými.
- 2) Průzkum korelační matice (k redukci pomocí PCA jsou zapotřebí silně korelované proměnné).
- 3) Provedení PCA a rozhodnutí o počtu hlavních komponent. Počet lze určit následovně:
 - Použití těch, které vysvětlují 90 % variability proměnných.
 - Použití těch, jejichž vlastní hodnota je vyšší než 1.
 - Použití těch, jež leží v klesající části grafu (angl. *scree plot*), který obrazuje závislost vysvětlené variability dat na počtu vybraných hlavních komponent.
- 4) Interpretace hlavních komponent.

Hlavní komponenty lze využít jako nezávisle proměnné pro regresní analýzu. Nevýhodou však může být, že PCA lze použít jen pro číselné proměnné. Dále je důležité si uvědomit, že výsledné komponenty jsou kombinací různých proměnných, ve výsledku jsou tudíž bezrozměrné, což může komplikovat jejich interpretaci.

2.3 Analýza závislostí

Pro analýzu závislostí jsou v metodice využity nástroje korelační a regresní analýzy:

- *Korelace* označuje vzájemný vztah mezi dvěma proměnnými. Pokud se jedna z nich mění, mění se korelativně i druhá a naopak. Sílu této závislosti vyjadřuje korelační koeficient.
- *Regresní analýza* je označení statistických metod, pomocí nichž lze odhadovat hodnoty závisle proměnné na základě znalosti jiných veličin (nezávisle proměnných). Je-li závislost prokázána, lze pomocí nezávisle proměnné odhadovat (predikovat) hodnoty závisle proměnné.

Zatímco korelační analýza odpovídá na otázku „Jak silná je závislost mezi proměnnými?“, regresní analýza poskytuje odpověď na otázku „Jak vypadá závislost mezi proměnnými?“. Důležité však je, že statistická závislost nemusí znamenat příčinnost (kauzalitu, tj. vztah mezi příčinou a následkem).

V různých kontextech se používá rozdílné názvosloví pro proměnné:

- závisle proměnná (Y): regresand, predikant, cílová proměnná, vysvětlovaná proměnná
- nezávisle proměnná (X): regresor, prediktor, predikční proměnná, kovariáta, vysvětlující proměnná

(označení Y, X , resp. y_i, x_i pro jednotlivé hodnoty, bude používáno i v dalším textu)

Je důležité si uvědomit, že hodnota korelačního koeficientu mezi proměnnými Y a X je stejná jako mezi proměnnými X a Y – z její velikosti tedy nelze rozlišit závisle a nezávisle proměnnou. Naopak u regrese Y na X nebo X na Y jsou výsledky rozdílné. Korelace by proto neměla být jediným nástrojem multifaktorové analýzy; měla by být spíše „přípravou“ na regresní analýzu.

2.3.1 Korelace

Jak bylo uvedeno, míru korelace určuje korelační koeficient; jeho hodnoty jsou v intervalu od -1 (antikorelace, nepřímá závislost) do 1 (úplná korelace, přímá závislost). Korelační koeficient může být jednoduchý nebo dílčí (parciální).

Používá se více druhů korelačních koeficientů, především v závislosti na typu analyzovaných dat (jejich přehled viz Obr. 4).



Obr. 4 Přehled druhů dat

Ordinální proměnná může být ordinální pořadová (vzniká seřazením jednotek změřením spojité proměnné) nebo ordinální kategoriální (rozeznává jen několik ordinálně uspořádaných kategorií, do kterých se jednotky zařazují). [10]

K analýze závislosti mezi nominálními proměnnými se používají kontingenční tabulky (např. koeficient Cramerovo V). Nejčastěji se však používají kvantitativní proměnné a k nim příslušné korelační koeficienty. Tab. 1 představuje doporučené druhy podle typu proměnných (P – Pearsonův korelační koeficient, S – Spearmanův korelační koeficient, K – Kendallův korelační koeficient).

Tab. 1 Přehled doporučených korelačních koeficientů v závislosti na druhu dat [10]

Druhá z proměnných	Jedna z proměnných			
	spojitá (normální)	spojitá (ne normální)	ordinální pořadová	ordinální kategoriální
spojitá (normální)	P			
spojitá (ne normální)	S	S		
ordinální pořadová	S	S	S	
ordinální kategoriální	K	K	K	K

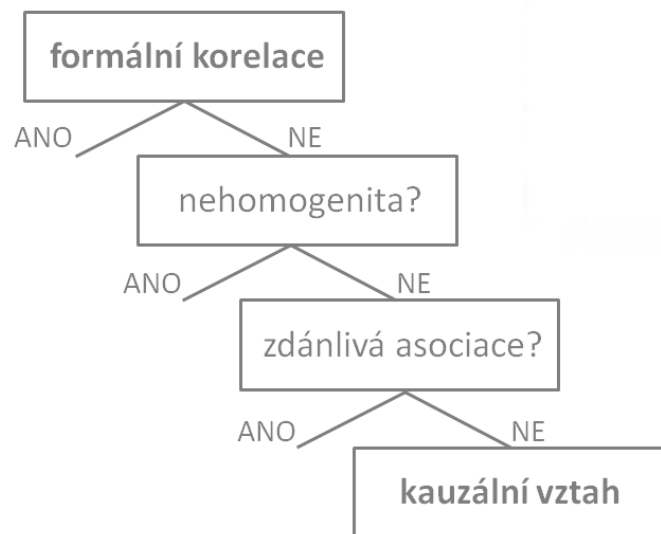
Spearmanův korelační koeficient (ρ) uvažuje pořadí hodnot jednotlivých proměnných a lze jej využít i u nelineárních vztahů.

Další druh korelačního koeficientu, tzv. Kendallovo tau (τ) vyjadřuje pravděpodobnost, že seřazení dvou náhodně vybraných jednotek podle proměnné X bude stejné jako podle proměnné Y .

Nejčastěji se používá Pearsonův korelační koeficient (r), který měří sílu vztahu dvou spojitých proměnných. Lze jej však použít jen u lineárních vztahů a proměnných s normálním rozdělením.

Obecně platí, že *jednoduchý korelační koeficient* se počítá mezi dvěma proměnnými: $r(Y, X)$ označuje korelační koeficient mezi Y a X .

Zároveň však platí, že korelace může být způsobena jinou (třetí) proměnnou, která koreluje s oběma studovanými proměnnými. Taková matoucí (rušivá) proměnná pak znesnadňuje interpretaci, protože nelze odlišit vliv matoucí a sledované proměnné. Lze uvést schéma pořadí, ve kterém by se měly nezajímavé korelace vyloučovat; výsledkem je korelace, která může indikovat kauzální vztah (viz Obr. 5). Jiní autoři uvádějí až devět kritérií [6]: za nejdůležitější se však označuje vyloučení zdánlivé asociace (působení jiné veličiny, tj. matoucí proměnné).



Obr. 5 Schéma kroků potřebných k prokázání kauzálního vztahu [10]

Podle velikosti korelačního koeficientu lze posuzovat sílu asociace. V různých oborech se udávají různé hodnoty, viz Tab. 2.

Tab. 2 Přehled síly závislosti mezi proměnnými podle velikosti korelačního koeficientu

Absolutní hodnota korelačního koeficientu	Síla závislosti
0,1 – 0,3	Malá
0,3 – 0,7	Střední
0,7 – 1,0	Velká

Výskyt takovýchto korelací lze však předpokládat spíše v experimentálním výzkumu. Při studiu dopravních jevů lze očekávat spolupůsobení mnoha proměnných a tudíž slabší korelační vztahy; některá doporučení proto považují korelační koeficienty za významné již od nižších hodnot, např. 0,4 – 0,5 [2].

Druhá mocnina korelačního koeficientu určuje tzv. *koeficient determinace*. Ten udává procento společné variability sledovaných proměnných, tedy kolik procent variability proměnné Y může být vysvětleno variabilitou proměnné X a naopak.

2.3.2 Lineární regrese

Klasický model lineární regrese Y na X je následující: $y_i = \beta_0 + \sum \beta_i x_i + \varepsilon_i$

kde:

β_0 regresní konstanta
 β_i regresní koeficienty
 ε_i rezidua

Koeficient determinace R^2 udává podíl vysvětlené variability na celkové variabilitě. Standardizovaný regresní koeficient β udává jednotkový přírůstek Y , když se X změní o jednotku (a ostatní se nemění).

Podstatou řešení regresní analýzy je: [12]

- 1) stanovení nejvhodnějšího tvaru regresního modelu (neboli regresní rovnice)
- 2) stanovení konkrétních hodnot parametrů (konstanty a koeficientů)
- 3) stanovení statistické významnosti modelu (tj. zda model podstatným způsobem přispěje ke zpřesnění odhadu závisle proměnné oproti výchozímu (nulovému) modelu)
- 4) interpretace výsledků daných modelem

Regresní analýza je velmi rozšířenou technikou statistické analýzy dat. Její standardní variantu však lze aplikovat pouze pro spojité proměnné, které (resp. jejichž rezidua) mají normální rozdělení. V ostatních případech lze použít tzv. zobecněnou lineární regresi (zobecněný lineární model), která bude uvedena dále.

2.3.3 Logistická regrese

Existuje mnoho situací, kdy závisle proměnná nabývá omezeného počtu hodnot – tento počet lze často redukovat na dvě hodnoty, které lze interpretovat jako výskyt ano/ne. Závisle proměnná je pak uvažována jako binární s hodnotami 1 (jestliže nastane sledovaný jev) nebo 0 (sledovaný jev nenastal). Cílem je zjistit, jak závisejí hodnoty pravděpodobnosti $P(Y = 1)$ na podmínkách, jež jsou dány hodnotami nezávislých proměnných. Hodnota této

pravděpodobnosti leží v intervalu mezi 0 a 1; pak platí, že pravděpodobnost $P(Y = 0) = 1 - P(Y = 1)$. Pro stručnost je dále označováno $P(Y = 1) = P$.

Šance (angl. *odds*) zkoumaného jevu je definována jako $\frac{P}{1-P}$, tj. poměr pravděpodobností existence a neexistence jevu (úspěchu a neúspěchu). Pro její logaritmus (tzv. *logit*) platí: $\ln\left(\frac{P}{1-P}\right) = \beta_0 + \sum \beta_i x_i$. Pravá strana rovnice udává průměrnou očekávanou hodnotu závisle proměnné, stejně jako v lineární regresi. Pravděpodobnost lze zpětně vypočítat pomocí tzv. logistické funkce: [10]

$$P = \frac{\exp(\beta_0 + \sum \beta_i x_i)}{1 + \exp(\beta_0 + \sum \beta_i x_i)}$$

Grafický průběh této funkce se nazývá logistická křivka (nebo také S-křivka). Uvedený model se nazývá logistický (nebo logitový) regresní model. Lze s ním, stejně jako pomocí lineárního regresního modelu, získat regresní (predikční) rovnice, jejichž regresní koeficienty udávají prediktivní schopnost použitých nezávislých proměnných.

Přestože logistická regrese je nejčastěji používána pro případy binární závisle proměnné, existují i další varianty pro nominální nebo ordinální proměnné. Pro odlišení se proto jednotlivé varianty označují jako binární, polytomická (multinomická) a ordinální logistická regrese.

K testu dobré shody regresního modelu s daty lze použít analogii R^2 Nagelkerka, jež lze hodnotit stejně jako při lineární regresi.

Regresní koeficient β_i lze interpretovat jako změnu logitu spojenou s jednotkovou změnou hodnoty nezávisle proměnné X_i za předpokladu, že hodnoty ostatních nezávisle proměnných se nezmění. Z výše uvedených rovnic vyplývá, že $\exp(\beta_i)$ lze interpretovat jako násobek analogické jednotkové změny šance. [17]

2.3.4 Zobecněný lineární model

Zobecněný lineární model (angl. *generalized linear model*, GLM) umožňuje modelovat závislost vysvětlované proměnné (kvantitativní nebo kvalitativní) na jednom nebo více prediktorech, které mohou opět být kvantitativní nebo kvalitativní. Důležitou předností GLM je, že umožňuje modelovat proměnné i s jinými pravděpodobnostními rozděleními, než je normální rozdělení použité v klasickém modelu regrese.

Předpokládá se, že systematická složka μ je prostřednictvím tzv. *spojovací funkce (link)* spojena s lineárním prediktorem η , což je lineární funkce parametrů modelu (systematická složka je funkcí lineárního prediktoru). V modelu se volí spojovací funkci g , vysvětlující proměnné X a na základě předpokladu o rozdělení vysvětlované proměnné Y se hledají takové koeficienty lineárního prediktoru, aby model co nejlépe vystihoval výsledky měření.

Například pro četnostní data (tj. diskrétní nezáporné hodnoty, což je i případ nehodovosti) se používá Poissonovo rozdělení, pro které se volí spojovací funkce $g(\mu) = \ln(\eta)$. Pro inverzní funkci pak platí $\mu = g^{-1}(\eta) = e^\eta$. [1]

Modelování se podle obecného postupu regrese z odstavce 2.3.2 provádí v následujících bodech:

1) *stanovení nejvhodnějšího tvaru regresního modelu*

Výchozí (nulový) model má tvar $Y = a$. V dalším kroku je vložena proměnná X , model má tvar $Y = a \cdot X^b$ (a je konstanta, b je koeficient). Model se vytváří krokově (*stepwise*) a to buď dopřednou selekcí (přidáváním proměnných, *forward stepwise*) nebo zpětnou eliminací (odebíráním proměnných, *backward stepwise*).

2) *stanovení konkrétních hodnot parametrů*

Zatímco u klasické lineární regrese se hodnoty parametrů určují nejčastěji metodou nejmenších čtverců (MNC), u GLM se nejčastěji používá metoda maximální věrohodnosti.

3) *stanovení statistické významnosti modelu*

Kritérium R^2 z klasické lineární regrese nelze u GLM použít. Lze využít tzv. informační kritéria, nejčastěji Akaikovo informační kritérium (AIC).

4) *interpretace výsledků daných modelem*

K hodnocení kvality modelu lze využít tzv. *věrohodnostní funkce* (angl. *likelihood*). Pro nulový model je věrohodnost minimální; pro další modely se hodnotí dvojnásobek rozdílu mezi jejich věrohodnostními funkcemi neboli *deviance*. Cílem je nalézt model, který zmenšuje celkovou devianci – ta je úměrná rozdílu logaritmu věrohodnostních funkcí (angl. *log-likelihood*, LL) mezi úplným modelem a nulovým modelem). Do modelu se postupně zařazují ty proměnné, které nejvíce snižují devianci vzhledem k aktuálnímu modelu. [15]

Jak bylo uvedeno výše, model lze vytvářet dopřednou selekcí nebo zpětnou eliminací; oba způsoby by měly ideálně vést ke stejnému výsledku. Výběr proměnných se může řídit dosaženou hladinou významnosti (signifikancí) – čím nižší, tím lepší. [4]

2.4 Aplikace v kontextu nehodovosti

Tato část navazuje na předchozí kapitolu a upřesňuje její aplikaci v kontextu nehodovosti. Konkrétní použití je předvedeno na dvou příkladech v následující praktické části metodiky.

V kontextu nehodovosti se používají víceproměnné regresní modely. Vysvětlovanou proměnnou bývá nehodovost (např. ve formě roční četnosti nehod na křižovatce nebo roční hustotě nehod na 1 km délky mezikřižovatkového úseku), vysvětlující proměnné se volí většinou podle dostupnosti: jedná se např. o intenzitu dopravy (RPDI), počet jízdních pruhů, šířku komunikace, rychlostní limit apod. Z toho vyplývá, že se většinou nejedná o *kauzální* modely (vysvětlující proměnné nerepresentují přímo nehodové faktory, jedná se spíše o proměnné *související* s nehodovostí).

Pro tyto modely se vžil označení *predikční* modely nehodovosti nebo nehodové predikční modely (angl. *accident prediction model*, APM nebo *crash prediction model*, CPM). Někdy se však uvádí, že označení predikční je zavádějící, protože použitá data reprezentují minulost a nelze tudíž zaručit jejich platnost v budoucnu (navíc při nesplněné kauzalitě); alternativní termín je *vysvětlující* modely [4, 15].

Modelování probíhá podle obecného postupu GLM uvedeném v odstavci 2.3.4. Pro zjednodušení zápisu budou místo označení β_i ($i = 0, 1, \dots$) použita písmena a, b, c, d . Používají se následující funkční tvary:

- na úsecích: $Y = a \cdot L^b \cdot I^c \cdot \exp(\sum d \cdot x)$
(tzn. je modelován vliv délky úseku L , intenzity I a dalších proměnných x),
 - na křižovatkách: $Y = a \cdot I_1^b \cdot I_2^c \cdot \exp(\sum d \cdot x)$
(tzn. je modelován vliv intenzity na hlavní komunikaci I_1 , intenzity na vedlejší komunikaci I_2 a dalších proměnných x),
- přičemž Y je nehodovost; a je regresní konstanta; b, c, d jsou regresní koeficienty.

Jako spojovací funkce zobecněného lineárního modelu se volí logaritmus. Jedná se o tzv. log-lineární multiplikativní model. Model odhaduje logaritmy hodnot $\ln(Y)$; hodnoty se získají prostřednictvím inverzní funkce $e^{\ln(Y)}$. [16] Např. výše uvedený model $Y = a \cdot L^b \cdot I^c \cdot \exp(\sum d \cdot x)$ se proto při výpočtu převádí na logaritmizovaný tvar $\ln(Y) = \ln(a) + b \cdot \ln(L) + c \cdot \ln(I) + \sum d \cdot x$.

U Poissonova rozdělení, zmíněného v odstavci 2.3.4, platí, že průměrná hodnota vysvětlované proměnné je rovna rozptylu. To však u četnosti nehod často neplatí: rozptyl zde často převyšuje průměrnou hodnotu. Z toho důvodu se používá varianta zvaná negativně binomické pravděpodobnostní rozdělení (NB). Jeho rozptyl je definován následovně:

$$\text{var}(Y) = \mu + k \cdot \mu^2$$

tj. rozptyl (variance) $\text{var}(Y)$ je roven průměru μ navýšenému o druhou část rovnice $k \cdot \mu^2$. Parametr k se nazývá *disperzní parametr* a jeho hodnota se získává v průběhu modelování. Lze jej mj. použít k hodnocení kvality modelu [13, 15].

3 Praktická část

V této části jsou uvedeny dva praktické příklady multifaktorové analýzy dopravní nehodovosti: na křižovatkách a na mezikřižovatkových úsecích. V příkladech byly aplikovány postupy popsané v teoretické části.

Ke zpracování byl využit statistický program IBM SPSS 19. Protože k tomuto software neexistuje česká lokalizace, uživatel musí pracovat v anglickém prostředí. Výstupy (tabulky a grafy) tuto skutečnost reflektují a jsou proto ponechány původní podobě; jsou však doplněny českými vysvětlivkami.

Příklad A – Křižovatky

Cílem tohoto příkladu bylo zjistit, jaké faktory zvyšují riziko nehody se zraněním na stykových křižovatkách v extravilánu.

A.1 Příprava dat

Byly vybrány stykové křižovatky v extravilánu na silnicích II. třídy v Jihomoravském kraji ($n = 72$). Závisle proměnnou je počet nehod se zraněním (N) v období 2007 – 2013 v okruhu 100 m od křižovatky (data získána od Policie ČR). Nezávisle proměnné vystihují expozici (intenzitu dopravy) a geometrické charakteristiky. Zdroje dat byly následující:

- Intenzita dopravy – z Celostátního sčítání dopravy 2010, na některých silnicích III. třídy a místních komunikacích doplněny vlastním krátkodobým průzkumem.
- Geometrické charakteristiky – ze Silniční databanky ŘSD ČR k lednu 2013, případně doplněny na základě mapových podkladů.

Intenzita dopravy byla vyčíslena jako počet vozidel vjíždějících do křižovatky (tj. polovina obousměrného RPD). Ve shodě se zahraničními zkušenostmi [18] byla intenzita uvažována v následujících tvarech:

- $F1$... intenzita na hlavní komunikaci [voz/den]
- $F2$... intenzita na vedlejší komunikaci [voz/den]
- FR ... podíl $F2/(F1 + F2)$
- FQ ... podíl $F2/F1$

Proměnné, popisující geometrii křižovatek, byly následující:

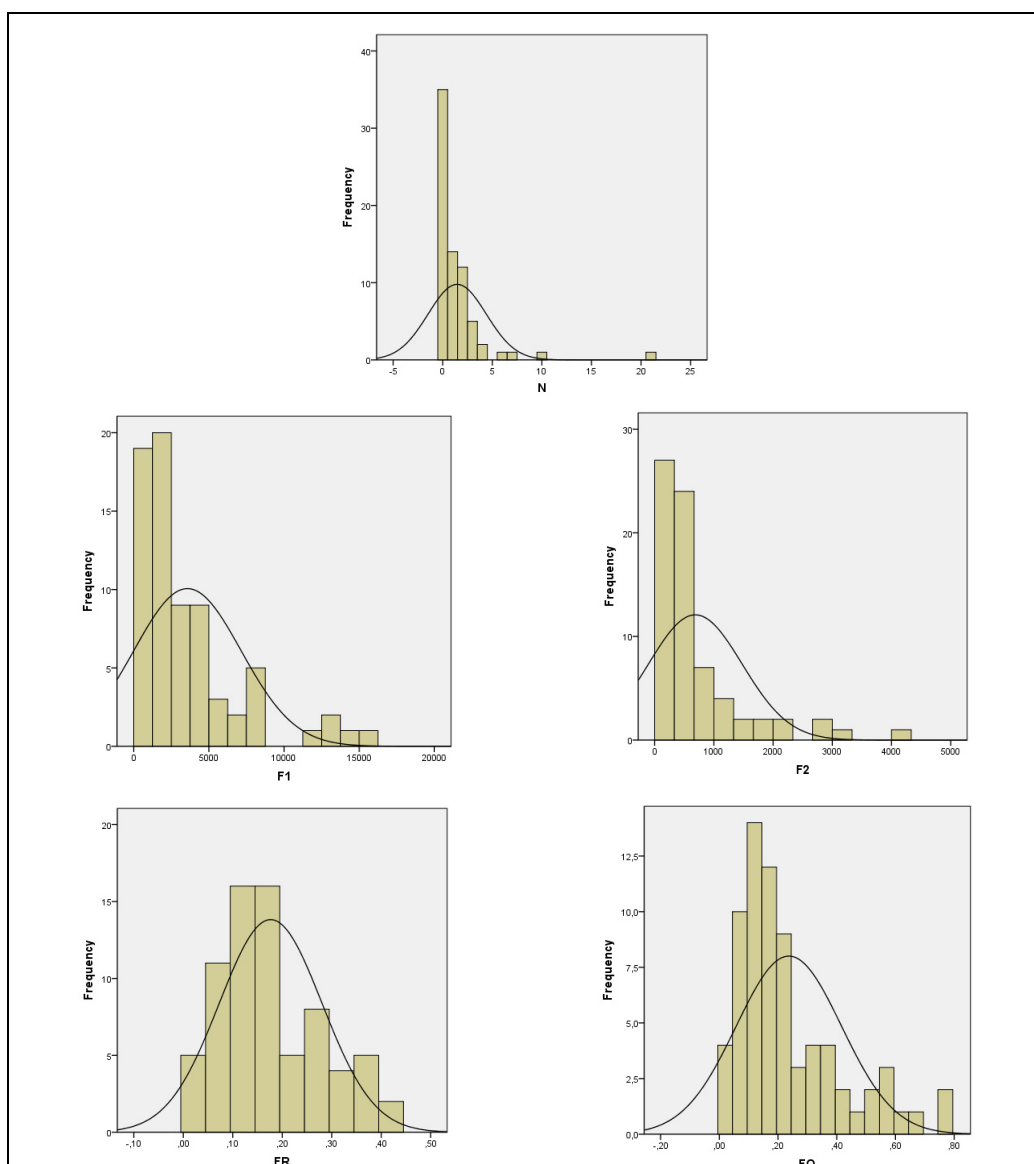
- ODB ... přítomnost samostatného odbočovacího pruhu na hlavní komunikaci (ano/ne)
- TYP_KRIZ ... typ křižovatky (prostá/usměrněná)
- $UHEL$... úhel křížení (vhodný/nevhodný)
- ZAL_PRED ... zalomená přednost (ano/ne)

A.2 Explorační analýza

Tabulka uvádí základní charakteristiky proměnných: u číselných proměnných minimum, maximum, průměr a směrodatnou odchylku; u kategoriálních proměnných četnost případů, označených 0 nebo 1.

proměnná	min.	max.	průměr	směr. odch.	četnosti
<i>N</i>	0	21	1,46	2,94	
<i>F1</i>	237	15022	3572,68	3569,44	
<i>F2</i>	9	4224	685,83	792,32	
<i>FR</i>	0,02	0,44	0,18	0,10	
<i>FQ</i>	0,02	0,79	0,24	0,18	
<i>ODB</i> <i>TYP_KRIZ</i> <i>UHEL</i> <i>ZAL_PRED</i>					0 (ne): 61 / 1 (ano): 11 0 (prostá): 45 / 1 (usměrněná): 27 0 (nevhodný): 13 / 1 (vhodný): 59 0 (ne): 57 / 1 (ano): 15

Dále jsou uvedeny histogramy číselných proměnných *N*, *F1*, *F2*, *FR*, *FQ*. U histogramů je zobrazena i křivka normálního rozdělení.



Z histogramů je zřejmé, že rozdělení hodnot proměnných pravděpodobně nepodléhá normálnímu rozdělení. Tento předpoklad lze statisticky testovat – nulová hypotéza je, že hodnoty pocházejí z normálního rozdělení. Na základě nízké dosažené hladiny významnosti (*Sig.*) lze nulovou hypotézu zamítnout, data nemají normální rozdělení.

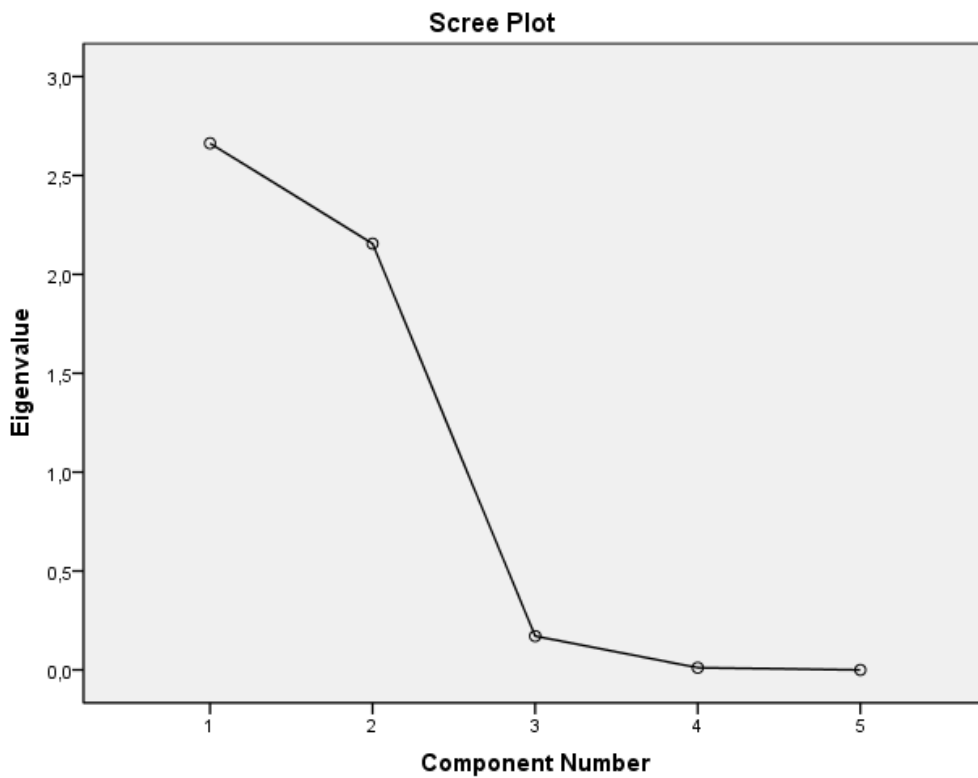
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
F1	,183	72	,000	,791	72	,000
F2	,233	72	,000	,719	72	,000
FR	,136	72	,002	,934	72	,001
FQ	,179	72	,000	,864	72	,000
N	,310	72	,000	,496	72	,000

a. Lilliefors Significance Correction

Platí tedy, že vysvětlovaná proměnná (nehodovost) i vysvětlující proměnné (intenzity) nemají normální rozdělení. Z grafů je navíc zřejmé, že závislosti nejsou lineární. Z těchto důvodů nelze použít lineární regresi.

V rámci explorační analýzy lze dále použít analýzu hlavních komponent (PCA). Lze ji ale použít jen na číselné proměnné. Výstupem PCA je následující graf.



Podle průběhu křivky jsou v datech přítomny 2 hlavní komponenty, viz tabulka:

Component Matrix

	Component	
	1	2
F1	,980	-,043
F2	,784	,529
FR	-,221	,970
FQ	-,232	,963

Podle největších faktorových zátěží (viz zvýrazněné hodnoty v tabulce) je zřejmé, že jednu komponentu tvoří *F1* a *F2* (intenzita na hlavní a vedlejší komunikaci), druhou komponentu tvoří *FR* a *FQ* (podíly intenzity na vedlejší komunikaci). Vzhledem k nízkému počtu proměnných i komponent se však redukce proměnných neprovede.

A.3 Analýza závislostí

A.3.1 Korelace

Protože proměnné nemají normální rozdělení a vztahy nejsou lineární, byla použita Spearmanova korelace (viz Tab. 1). Tabulka uvádí hodnoty korelačních koeficientů (pro přehlednost jsou vynechány hodnoty pod hlavní diagonálou).

Correlations

	F1	F2	FR	FQ
F1	1,000	,715	-,316	-,310
F2		1,000	,377	,383
FR			1,000	,999
FQ				1,000

Zvýrazněné jsou hodnoty významných (a signifikantních) korelací: zaprvé mezi *F1* a *F2*, zadruhé mezi *FR* a *FQ*. To potvrzuje i zjištění předchozí analýzy hlavních komponent.

V dalších krocích by tedy bylo vhodné využít vždy jen jednu proměnnou z obou skupin. Toto zjištění bude potvrzeno v pozdější analýze GLM (odstavec A.3.3).

A.3.2 Logistická regrese

Vysvětlovanou proměnnou byl počet nehod, převedený na binární proměnnou (hodnota „0“ pro 0 nehod, hodnota „1“ pro počet nehod > 0). Jako vysvětlující proměnné byly do regresního modelu dosazeny všechny ostatní proměnné. Podle jejich dosažené hladiny významnosti (*Sig.*) byly postupně vylučovány; výsledný model obsahuje *F1*, *F2* a *ODB* (viz tabulka).

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
F1	,001	,000	11,070	1	,001	1,001
F2	,002	,001	3,379	1	,066	1,002
ODB(1)	4,656	1,708	7,433	1	,006	105,247
Constant	-7,280	2,188	11,070	1	,001	,001

Vysvětlivky: *B* – regresní koeficienty, *S.E.* – standardní chyba, *Wald* – Waldovo testové kritérium, *df* – stupně volnosti, *Sig.* – dosažená hladina významnosti. Poslední sloupec *Exp(B)* uvádí vliv regresních koeficientů (šanci), viz zvýrazněné hodnoty. Jejich interpretaci lze provádět např. následovně:

- Když se intenzita na hlavní komunikaci (*F1*) zvýší o 1, šance na nehodu v křižovatce vzroste 1,001×. To je nízké číslo, je však nutno si uvědomit příslušné jednotky [voz/den]. V případě uzavírky a převedení dopravy se může intenzita zvýšit např. o 3000 voz/den, v tom případě se 1,001 umocní na 3000, tj. šance vzroste cca 20×.
- Není-li na křižovatce samostatný odbočovací pruh, šance na nehodu (ve srovnání s variantou se samostatným odbočovacím pruhem) vzroste cca 105×.

Obě zjištění jsou v souladu se zkušeností: intenzita je hlavním rizikovým faktorem, usměrnění odbočujícího proudu do samostatného pruhu je považováno za bezpečné.

A.3.3 Zobecněný lineární model

Je známo, že směrodatná odchylka je rovna druhé odmocnině z rozptylu. Podle hodnoty směrodatné odchylky počtu nehod 2,94 v tabulce v části A.2 je tedy rozptyl $2,94^2 = 8,64$. Tato hodnota je výrazně vyšší než průměr (1,46), v GLM tedy bylo zvoleno negativně binomické rozdělení.

Podobně jako u logistické regrese se model vytváří postupně; nejprve se všemi proměnnými, které se postupně redukuje tak, aby zůstaly jen ty se statisticky významným vlivem (*Sig.*). Výsledný model obsahuje proměnné $\ln(F1)$, $\ln(FR)$ a *ODB* (viz tab.)

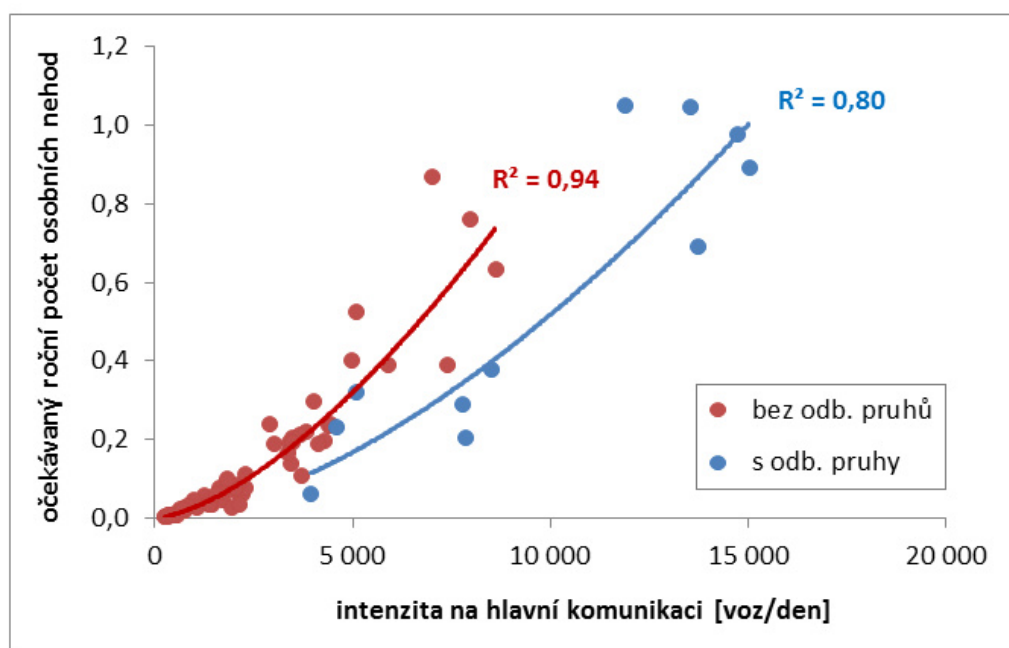
Parameter Estimates

Parameter	B	S.E.	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-12,503	2,1723	-16,761	-8,245	33,128	1	,000
[ODB=0]	,616	,3788	-,127	1,358	2,643	1	,104
[ODB=1]	0
ln_F1	1,625	,2393	1,156	2,094	46,075	1	,000
ln_FR	,553	,2480	,067	1,039	4,969	1	,026
(Scale)	1						
(Negative binomial)	,245	,1327	,085	,708			

Vysvětlivky: B – regresní koeficienty, $S.E.$ – standardní chyba, $Wald$ – Waldův interval spolehlivosti a testové kritérium, df – stupně volnosti, $Sig.$ – dosažená hladina významnosti. Jako „negative binomial“ je označen disperzní parametr modelu.

Míra vlivu jednotlivých vysvětlujících proměnných je zřejmá z velikosti Waldova testového kritéria (zvýrazněné hodnoty): nejvýznamnější je vliv $F1$, dále FR a ODB .

Jedna z možností, jak tyto vlivy vyjádřit graficky, je na grafu (Obr. 6): roční očekávaná nehodovost je modelována v závislosti na intenzitě na hlavní komunikaci ($F1$) pro dvě varianty ODB (bez odbočovacích pruhů, s odbočovacími pruhy). Z průběhu křivek (mocninných trendů) je zřejmý vliv přítomnosti odbočovacích pruhů: bez nich (červená křivka) lze očekávat obecně vyšší nehodovost než s nimi (modrá křivka).



Obr. 6 Graf očekávané nehodovosti v závislosti na intenzitě dopravy a přítomnosti odbočovacích pruhů

Příklad B – Mezikřižovatkové úseky

V tomto příkladu byly identifikovány charakteristiky silniční sítě, které zvyšují riziko dopravní nehody bez ohledu na její příčinu či závažnost.

B.1 Příprava dat

Byly vybrány extravilánové úseky na silnicích II. třídy v Jihomoravském kraji, vymezené mezi křižovatkami (celková délka 995 km). Protože se u těchto úseků průběžně mění jejich vlastnosti, byla nejprve provedena segmentace do homogenních úseků. Homogenita byla definována ve smyslu vybraných charakteristik představujících potenciální rizikové faktory, které mají v rámci úseků stejnou hodnotu. Tyto faktory zahrnovaly intenzitu, nejvyšší

dovolenou rychlost, šířkovou kategorii komunikace, počet jízdních pruhů a přítomnost zpevněné krajnice. Při každé změně hodnoty některé z těchto veličin byl vytvořen nový úsek.

Aby byly délky výsledných úseků vhodné pro případné prohlídky na místě, byla segmentace provedena tak, aby jejich maximální délka nepřesahovala 500 m. Delší úseky byly rozděleny na 250m části. Výsledkem segmentace bylo 3764 úseků s homogenními hodnotami zmíněných charakteristik, z toho bylo 83 % úseků o délce ≤ 250 m a 17 % úseků o délce > 250 m.

Ke každému úseku byla přiřazena nehodová, silniční a dopravní data z následujících zdrojů:

- Počet všech nehod zaznamenaných Policií ČR za období 2009 – 2012 (N).
- Údaje o silniční síti ze Silniční databanky ŘSD ČR (aktualizace k červenci 2010), případně doplněné z mapových podkladů.
- Údaje o intenzitě dopravy ze Silniční databanky ŘSD ČR (Celostátní sčítání dopravy 2010).
- Údaje o kvalitě povrchu vozovky z průzkumu firmy PavEx Consulting v roce 2011.

Z uvedených dat byly vytvořeny následující proměnné:

- expozice: intenzita (celkové RPDI a podíl těžkých nákladních vozidel), délka
- kategorie komunikace (S 7,5; 9,5; 11,5), počet jízdních pruhů, přítomnost zpevněné krajnice (ano/ne), nejvyšší dovolená rychlost jízdy (nesnížená/snížená), hustota křižovatek a dopravních zařízení (počet na 1 km)
- křivolakost trasy ve 3 formách (průměrná, závlek, hustota oblouků, viz [3]), přítomnost lesa v okolí komunikace (ano/ne)
- podíl délky úseku, která má havarijní stav povrchu vozovky (číslo v intervalu 0 – 1)

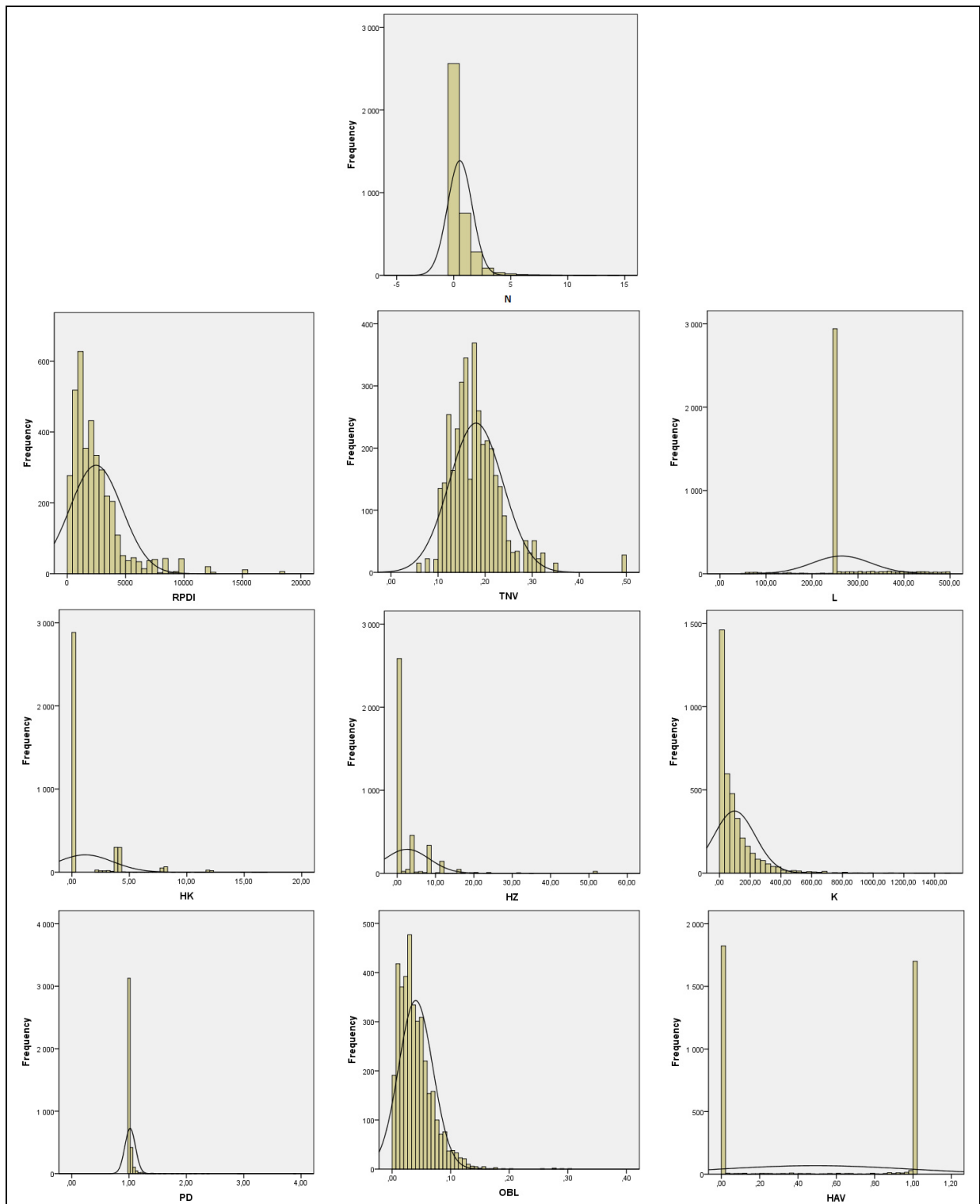
B.2 Explorační analýza

Charakteristiky proměnných jsou uvedeny v tabulce (u číselných proměnných minimum, maximum, průměr, směrodatná odchylka; u kategoriických proměnných četnosti):

proměnná	popis	min.	max.	průměr	směr. odch.	četnosti
<i>N</i>	počet nehod (2009-2012)	0	14	0,54	1,08	
<i>RPDI</i>	roční průměrná denní intenzita dopravy (RPDI) [voz/den]	91	18498	2459,37	2230,15	
<i>TNV</i>	podíl intenzity těžkých nákladních vozidel	0,06	0,50	0,18	0,06	
<i>L</i>	délka [m]	51,00	499,88	264,29	64,03	
<i>HK</i>	hustota křižovatek [km^{-1}]	0,00	16,90	1,16	2,40	
<i>HZ</i>	hustota dopravních zařízení [km^{-1}]	0,00	52,00	2,58	5,76	
<i>K</i>	průměrná křivolakost trasy [gon/km]	0,00	1498,18	98,03	134,27	
<i>PD</i>	podíl délky úseku vůči délce přímé spojnice	1,00	3,90	1,02	0,09	
<i>OBL</i>	hustota oblouků	0,00	0,30	0,04	0,03	
<i>HAV</i>	podíl délky úseku s havarijním stavem povrchu vozovky	0,00	1,00	0,49	0,49	
<i>KAT</i>	kategorie komunikace					0 (S 7,5): 3156 / 1 (S 9,5 nebo S 11,5): 608
<i>PRUH</i>	počet jízdních pruhů					0 (2 pruhy): 3622 / 1 (více pruhů): 142
<i>KRAJ</i>	přítomnost zpevněné krajnice					0 (ne): 3333 / 1 (ano): 431
<i>RYCH</i>	snížení nejvyšší dovolené rychlosti jízdy [km/h]					0 (ne): 3722 / 1 (ano): 42
<i>LES</i>	přítomnost lesa v okolí komunikace					0 (ne): 2977 / 1 (ano): 787

Podle četností je zřejmé, že hodnoty proměnných *PRUH* (počet jízdních pruhů) a *RYCH* (snížení nejvyšší dovolené rychlosti jízdy) mají malou variabilitu: 95% úseků bylo dvoupruhových a 99% úseků nemělo snížený rychlostní limit. Obě proměnné tedy byly vyloučeny.

Dále byly vytvořeny histogramy číselných proměnných. U histogramů je zobrazena i křivka normálního rozdělení.



Z histogramů je zřejmé, že rozdělení hodnot proměnných pravděpodobně nepodléhá normálnímu rozdělení. Tento předpoklad lze statisticky testovat – nulová hypotéza je, že hodnoty pocházejí z normálního rozdělení. Na základě nízké dosažené hladiny významnosti (*Sig.*) lze nulovou hypotézu zamítnout, data nemají normální rozdělení.

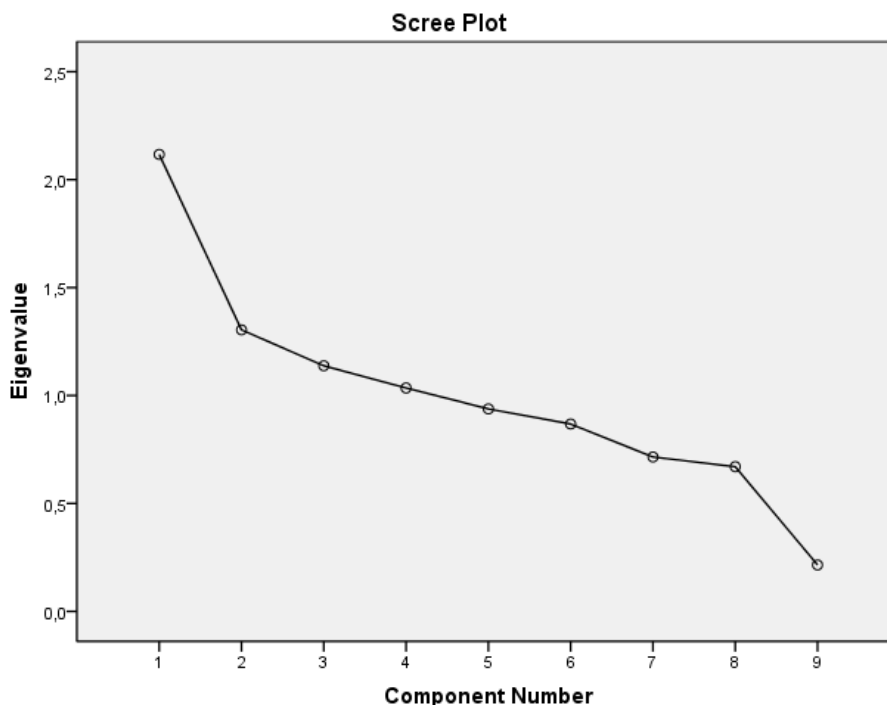
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
N	,371	3764	,000	,546	3764	,000
RPDI	,155	3764	,000	,772	3764	,000
TNV	,084	3764	,000	,901	3764	,000
L	,417	3764	,000	,604	3764	,000
HK	,452	3764	,000	,540	3764	,000
HZ	,360	3764	,000	,484	3764	,000
K	,233	3764	,000	,681	3764	,000
PD	,416	3764	,000	,166	3764	,000
OBL	,106	3764	,000	,884	3764	,000
HAV	,317	3764	,000	,659	3764	,000

a. Lilliefors Significance Correction

Platí tedy, že vysvětlovaná proměnná (nehodovost) i vysvětlující proměnné nemají normální rozdělení. Z grafů je navíc zřejmé, že závislosti nejsou lineární. Z těchto důvodů nelze použít lineární regresi.

V rámci explorační analýzy lze dále použít analýzu hlavních komponent (PCA). Lze ji ale použít jen na číselné proměnné. Výstupem PCA je následující graf.



Z průběhu křivky je zřejmá pouze jedna výrazná hlavní komponenta. To potvrzuje i tabulka, kde nejvýraznější faktorové zátěže tvoří proměnné *K*, *PD*, *OBL* – všechny vystihují křivolakost trasy.

Component Matrix

	Component			
	1	2	3	4
RPDI	-,117	,643	-,441	-,084
TNV	-,138	-,575	,509	,042
HK	,006	,523	,456	-,040
HZ	,091	,432	,535	,254
L	,028	-,205	-,421	,590
K	,930	-,076	,010	-,044
PD	,770	-,109	-,052	,004
OBL	,779	,127	,001	-,107
HAV	,101	,156	,095	,774

Z průběhu grafu je však zřejmé, že důležitost hlavních komponent není striktně oddělena. Nebude tedy redukován počet proměnných, ale budou dále použity v původním počtu.

B.3 Analýza závislostí

B.3.1 Korelace

Protože proměnné nemají normální rozdělení a vztahy nejsou lineární, byla použita Spearmanova korelace (viz Tab. 1). Tabulka uvádí příslušné korelační koeficienty (pro přehlednost byla vynechána část matice pod hlavní diagonálou).

Correlations

	RPDI	TNV	HK	HZ	L	K	PD	OBL	HAV
RPDI	1,000	-,248	,126	,043	-,012	-,217	-,215	-,010	-,005
TNV		1,000	-,036	,001	-,014	-,092	-,083	-,137	-,034
HK			1,000	,242	-,053	-,030	-,031	,006	,001
HZ				1,000	-,039	,012	,004	,033	,047
L					1,000	,054	,129	,002	,022
K						1,000	,961	,715	,061
PD							1,000	,670	,064
OBL								1,000	,047
HAV									1,000

Zvýrazněné koeficienty ukazují na silnou korelaci mezi proměnnými *K*, *PD* a *OBL*. *K* bylo ponecháno, ostatní (jako odvozené veličiny) byly vyloučeny.

B.3.2 Logistická regrese

Počet nehod byl převeden na binární proměnnou (hodnota „0“ pro 0 nehod, hodnota „1“ pro počet nehod > 0). Jako vysvětlující proměnné byly do regresního modelu dosazeny

všechny ostatní proměnné. Podle jejich dosažené hladiny významnosti (*Sig.*) byly postupně vylučovány; výsledný model obsahuje proměnné *RPDI*, *TNV*, *LES*, *L*, *K* (viz tabulka).

	B	S.E.	Wald	df	Sig.	Exp(B)
RPDI	,000	,000	188,963	1	,000	1,000
TNV	-3,356	,741	20,519	1	,000	,035
LES(1)	-,341	,090	14,346	1	,000	,711
L	,004	,001	51,193	1	,000	1,004
K	,001	,000	12,808	1	,000	1,001
Constant	-1,720	,229	56,489	1	,000	,179

Vysvětlivky: *B* – regresní koeficienty, *S.E.* – standardní chyba, *Wald* – Waldovo testové kritérium, *df* – stupně volnosti, *Sig.* – dosažená hladina významnosti. Poslední sloupec *Exp(B)* uvádí vliv regresních koeficientů (šanci), viz zvýrazněné hodnoty. Jejich interpretaci lze provádět např. následovně:

- Když se zvýší podíl těžké dopravy (*TNV*) z 0,1 na 0,3 (změna o 0,2), šance na nehodu se sníží hodnotou $0,035^{0,2}$, tj. cca 0,5× (sníží se na polovinu). Důvodem může být snížení rychlosti.
- Na úseku, kolem kterého není les (*LES* = 0), je cca 0,7× nižší šance nehody (ve srovnání s úsekem, kolem kterého je les). To potvrzuje skutečnost, že lesní prostředí může souviset s mokrou vozovkou, rozdíly v osvětlení apod.

B.3.3 Zobecněný lineární model

Je známo, že směrodatná odchylka je rovna druhé odmocnině z rozptylu. Podle hodnoty směrodatné odchylky počtu nehod 1,08 v tabulce v části B.2 je tedy rozptyl $1,08^2 = 1,17$. Tato hodnota je vyšší než průměr (0,54), v GLM tedy bylo zvoleno negativně binomické rozdělení. Podle zahraničních zkušeností [9] byla intenzita dosazena ve dvou formách: *RPDI* a $\ln(RPDI)$.

Podobně jako u logistické regrese se model vytváří postupně. Proměnné byly vylučovány podle hladiny významnosti (*Sig.*). Proměnné ve výsledném modelu jsou patrné z následující tabulky.

Parameter Estimates

Parameter	B	S.E.	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
			(Intercept)	-10,302			
ln_RPDI	1,189	,0817	1,029	1,349	211,967	1	,000
RPDI	-7,664E-5	2,2225E-5	,000	-3,308E-5	11,892	1	,001
L	,003	,0004	,002	,003	41,354	1	,000
HZ	-,014	,0055	-,025	-,003	6,583	1	,010
K	,002	,0002	,001	,002	68,847	1	,000
[KRAJ=0]	,263	,0850	,096	,430	9,568	1	,002
[KRAJ=1]	0
[LES=0]	-,482	,0689	-,617	-,347	48,944	1	,000
[LES=1]	0
(Scale)	1						
(Negative binomial)	,774	,0713	,646	,927			

Vysvětlivky: *B* – regresní koeficienty, *S.E.* – standardní chyba, *Wald* – Waldův interval spolehlivosti a testové kritérium, *df* – stupně volnosti, *Sig.* – dosažená hladina významnosti. Jako „negative binomial“ je označen disperzní parametr modelu.

Podle velikosti Waldova testového kritéria (zvýrazněné hodnoty) jsou nejvýraznější vlivy proměnných $\ln(RPDI)$, K , LES , L . Riziko se zvyšuje s délkou úseku, což je logické. Přítomnost lesa v okolí komunikace rovněž zvyšuje riziko dopravní nehody, to může být způsobeno obtížnějšími světelnými poměry, přítomností nečistot na povrchu vozovky, delší dobou schnutí po dešti, vyšším rizikem námrazy apod.

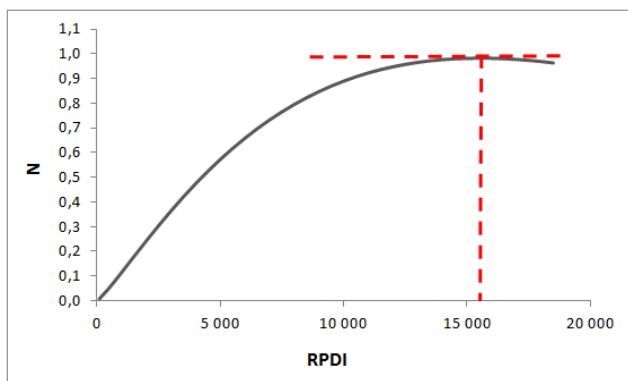
Vliv proměnných lze zobrazit i graficky. Pro ilustraci jsou zobrazeny závislosti očekávané nehodovosti N za čtyřleté období (2009 – 2012) na intenzitě ($RPDI$) a křivolakosti (K). Využita je rovnice modelu vytvořená pomocí koeficientů ve sloupci *B*.

$$N = e^{-10,302} \cdot RPDI^{1,189} \cdot e^{-7,664 \cdot 10^{-5} \cdot RPDI} \cdot e^{0,003 \cdot L} \cdot e^{-0,014 \cdot HZ} \cdot e^{0,002 \cdot K} \cdot \begin{cases} e^{0,263} & \text{pro } KRAJ = 0 \\ 1 & \text{pro } KRAJ = 1 \end{cases} \cdot \begin{cases} e^{-0,482} & \text{pro } LES = 0 \\ 1 & \text{pro } LES = 1 \end{cases}$$

V rovnici jsou vyznačeny dílčí vlivy intenzity a křivolakosti; po vynásobení konstantou se získají dílčí rovnice:

- vliv intenzity: $e^{-10,302} \cdot RPDI^{1,189} \cdot e^{-7,664 \cdot 10^{-5} \cdot RPDI}$
- vliv křivolakosti: $e^{-10,302} \cdot e^{0,002 \cdot K}$

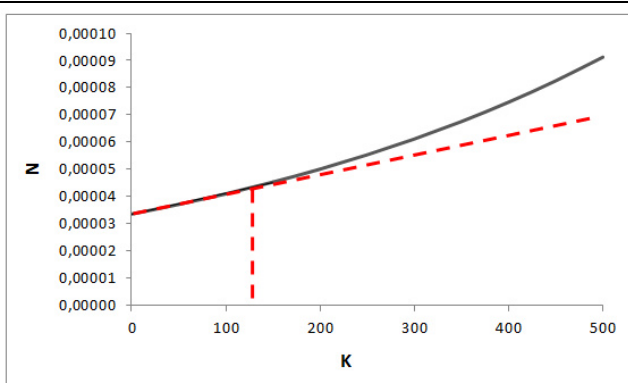
Z těchto rovnic byl odvozen trend, použitý v následujících grafech.



Obr. 7 Závislost očekávané nehodovosti na RPDÍ

Vliv intenzity:

Nehodovost má rostoucí trend přibližně do hodnoty intenzity 15 500 voz/den, následuje pokles. „Zlomová“ hodnota představuje hranici kapacity, při překročení její hodnoty dochází ke kongescím, tj. snížení rychlosti a tudíž i nehodovosti. Tato hodnota přibližně odpovídá mezní kapacitě silnic II. třídy uvedené v ČSN 73 6101, tab. 5.



Obr. 8 Závislost očekávané nehodovosti na křivolakosti úseku

Vliv křivolakosti:

Nehodovost má rostoucí trend, nejprve lineární, prudší nárůst nastává přibližně od hodnoty křivolakosti 125 gon/km. Tento údaj odpovídá zahraničním zkušenostem [8], podle kterých má křivolakost vliv na nehodovost od hodnot 100 – 150 gon/km. Důvodem mohou být zhoršené rozhledové poměry nebo změny jízdní rychlosti.

Závěr

Pro ilustraci možnosti praktického postupu multifaktorové analýzy dopravní nehodovosti byly uvedeny příklady dvou prvků silniční sítě: křižovatky a mezikřižovatkové úseky. V obou případech byla využita data Policie ČR, dále Celostátní sčítání dopravy a Silniční databanka ŘSD ČR; bylo však nutno je doplnit i o vlastní průzkum, např. doplnění intenzity na nesčítaných komunikacích nebo další údaje z mapových podkladů.

Byla ilustrována důležitost explorační analýzy (seznámení se s daty). Při této analýze bylo např. zjištěno, že hodnoty některých proměnných nemají nedostatečnou variabilitu, nejsou tudíž vhodné pro regresní analýzu. Dále bylo zjištěno, že proměnné nemají normální rozdělení a lineární závislosti, je tedy zřejmé, že nelze použít klasickou regresi.

Následující analýza hlavních komponent se používá spíše v případech velkého množství proměnných. Protože to v příkladech neplatilo (metodu navíc nelze aplikovat na kategorické proměnné), zjištěných komponent bylo relativně málo. Závěry analýzy tedy nebyly využity a původní proměnné byly ponechány.

Důležitá byla korelační analýza – byla-li zjištěna silná závislost mezi vysvětlujícími proměnnými, bylo nutno některou vyloučit.

Na závěr byly provedeny dvě analýzy regrese – logistická regrese a zobecněná regrese (zobecněný lineární model). Z jejich výsledků lze interpretovat míru vlivu, jaký mají jednotlivé proměnné na nehodovost.

Prostřednictvím provedených analýz byly identifikovány následující rizikové faktory:

	Křižovatky (příklad A)	Úseky (příklad B)
Analýza hlavních komponent	intenzity	křivolakost
Logistická regrese	intenzity, sam. odb. pruh	intenzita, křivolakost, les...
Zobecněný lineární model	intenzity, sam. odb. pruh	intenzita, křivolakost, les...

Závěry jsou silnější v případě, kdy je vliv některého faktoru potvrzen více analýzami. Takto byl potvrzen vliv přítomnosti samostatného odbočovacího pruhu u křižovatek, dále křivolakosti a lesa u mezikřižovatkových úseků.

Závěry jsou však vždy omezeny dostupností použitých dat. Zároveň platí, že řada veličin, které by teoreticky měly být rizikovými faktory, nemusí mít statisticky prokazatelný vliv a do modelu tudíž nejsou zahrnuty – u křižovatek se např. neuplatnil úhel křížení, u modelu úseků se neuplatnila proměnná popisující havarijní stav povrchu vozovky.

Závěrem byly některé zjištěné vlivy ilustrovány graficky. U křižovatek byl takto prokázán vliv samostatných odbočovacích pruhů; u úseků byl ukázán vliv intenzity a křivolakosti, jenž i číselně koresponduje s jinými zdroji.

Příklady ilustrují, že:

- Každý způsob analýzy poskytuje odlišný pohled, je proto vhodná jejich kombinace.
- Vypovídací schopnost je dána kvalitou modelů, jež jsou omezeny dostupností použitých dat. Nejčastěji se používají plošné databáze, je však nutno je doplňovat vlastními průzkumy.

4 Přínosy metodiky

4.1 Novost postupu

Postupy, uvedené v metodice, se používají se v jiných vědních oborech, inovativní je však jejich použití v kontextu dopravní nehodovosti. Představené analýzy umožňují identifikovat a kvantifikovat rizikové faktory v rámci větší silniční sítě.

Výsledné zobecněné lineární modely mají zároveň další možné aplikace a to podle zaměření:

- zaměření na soubor jednotek (např. soubor křižovatek dané konfigurace v daném kraji) ... zde se využije obecná rovnice modelu
- zaměření na jednotku (např. konkrétní křižovatku z uvedeného souboru) ... zde se do rovnice dosadí konkrétní hodnoty a zpřesní se podle empirické bayesovské metody

Tab. 3 uvádí otázky, na které mohou tyto dvě zaměření odpovědět, spolu s konkrétními příklady.

Tab. 3 Otázky a příklady demonstrující dvě různá zaměření multifaktorových analýz [19]

Zaměření na soubor	Zaměření na jednotku
Jaká je „normální“ bezpečnost jednotek daných vlastností? (Příklad: <i>Jaká je nehodovost vybrané kategorie křižovatek?</i>)	Je tato jednotka „abnormálně“ nebezpečná? (Příklad: <i>Jedná se o kritické místo?</i>)
Jak se liší bezpečnost dvou souborů? (Příklad: <i>Jsou směrově dělené komunikace bezpečnější než směrově nedělené?</i>)	Jaký bude přínos změny vlastností konkrétní jednotky? (Příklad: <i>Jak se změní nehodovost průsečné křižovatky po přestavbě na okružní?</i>)

Ve srovnání se standardními postupy mají metody doporučené v metodice několik předností, které jsou pro přehlednost uvedeny v Tab. 4 (rozděleny podle zaměření uvedeného v Tab. 3).

Tab. 4 Srovnání standardních a doporučených metod

Zaměření	Otázky	Standardní metody	Metody doporučené v metodice
Soubor	Jsou metody vhodné pro analýzu nehodovosti (četnosti nehod)?	Standardní metody (např. lineární regresi) lze aplikovat pouze na spojitá data s normálním pravděpodobnostním rozložením. Četnost nehod těmto požadavkům neodpovídá.	Ano, zobecněný lineární model zohledňuje statistické vlastnosti četnostních dat (diskrétní data s negativně binomickým pravděpodobnostním rozdělením).
	Kolik současně působících vlivů lze posuzovat?	Vlivy lze posuzovat (tabelovat nebo zobrazovat v grafu) vždy maximálně dva zároveň.	Lze analyzovat neomezený počet současně působících vlivů.
Jednotka	Jsou identifikována místa, která jsou svým uspořádáním skutečně rizikovější?	Standardní metody neumožňují redukci náhodné složky ve výskytu nehod. Velký podíl míst je tudíž identifikován nesprávně.	Použitím empirického bayesovského přístupu lze redukovat vliv náhodné složky. Podíl nesprávně identifikovaných míst je tudíž výrazně nižší než u standardních metod.
	Jsou kritická místa identifikována shodně v různých časových obdobích?	Vlivem působení náhodné složky se místa identifikovaná v různých obdobích výrazně liší.	Velká část identifikovaných míst je v čase stabilní.

Z tabulky vyplývá, že standardní metody nejsou pro analýzu multifaktorových jevů vhodné; navíc je nelze použít s ohledem na statistické vlastnosti nehodových dat. Řešením jsou proto postupy, představené v této metodice. Pomocí představeného zobecněného lineárního modelu lze kvantifikovat vliv jednotlivých spolupůsobících proměnných a to samostatně (bez ovlivnění ostatními proměnnými). Konkrétně tedy např. graf na Obr. 7 ukazuje jak je modelovaná nehodovost ovlivněna intenzitou (s odfiltrováním vlivu délky úseku, křivolakosti ad.). Oproti tomu standardními metodami nelze tyto vlivy odfiltrovat, výsledek je vždy ovlivněn všemi proměnnými.

4.2 Popis uplatnění certifikované metodiky

Metodiku lze uplatnit jako návod k provádění multifaktorové analýzy dopravní nehodovosti. Analýza umožňuje identifikovat a kvantifikovat rizikové faktory v rámci větší silniční sítě. Může být tedy součástí procesu řízení bezpečnosti silniční sítě.

Analýza vychází z dat o souboru prvků silniční sítě, nezabývá se tedy pouze jednotlivými místy, ale její výsledky lze v závěru na konkrétní místa aplikovat. Na jednom úseku například nelze zjišťovat vliv hodnot křivolakosti na bezpečnost (protože úsek má jen jednu hodnotu křivolakosti). Proto se analyzuje soubor úseků a z výsledku se zjistí vliv křivolakosti na dotyčných úsecích. Podle kritické hodnoty (podle grafu na Obr. 8 je to křivolakost 125

gon/km) pak lze na konkrétním místě hodnotit, zda je křivolakost bezpečná či nebezpečná a zda nepřispívá či přispívá k nehodovosti.

Lze uvést konkrétní aplikace, kdy se postupy prezentované v metodice, uplatnily ve studiích CDV, v souladu s dělením v Tab. 3:

- Z pohledu celého souboru byly ve studiích CDV srovnávány např. stykové a průsečné křižovatky v Praze [29, 30] nebo směrově dělené a nedělené úseky sítě TEN-T [32].
- Z pohledu jednotek se jednalo o úlohy identifikace kritických míst [20 – 27]. Lze uvést např. identifikaci kritických míst v extravilánu v Jihomoravském kraji na silnicích II. třídy [31, 33, 34, 37] nebo na silnicích I. třídy [35]. Ohledně dopravně bezpečnostních opatření se hodnotila např. účinnost přestaveb průsečných křižovatek na okružní [36].

Z uvedeného výčtu studií vyplývá, že metodu lze aplikovat v intravilánu i extravilánu; na křižovatkách i úsecích; na městské, krajské i celostátní silniční síti.

4.3 Ekonomické aspekty

Jak je uvedeno v předchozím odstavci, modely získané multifaktorovou analýzou prováděné podle předložené metodiky, mají řadu uplatnění s ekonomickým aspektem. Mohou pomoci při rozhodování o tom

- na co by se měla opatření přednostně zaměřit (na základě kvantifikace kritických faktorů),
- kde by se měla opatření přednostně aplikovat (na základě identifikace kritických míst),
- jaká opatření jsou nejpřínosnější (na základě hodnocení účinnosti).

Uvedené racionalizace postupu mohou vést k ekonomickým úsporám na straně správce příslušné silniční sítě. Výsledné zvýšení bezpečnosti a snížení nehodovosti vede zároveň ke snížení celospolečenských socioekonomických ztrát.

5 Reference

5.1 Seznam použité související literatury

V tomto seznamu jsou uvedeny zdroje, které byly v metodice použity.

- [1] Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., Thandi, N. (2008). *A Practitioner's Guide to Generalized Linear Models*. CAS, Arlington.
- [2] Council, F.M., Reinfurt, D.W., Campbell, B.J., Roediger, F.L., Carroll, C.L., Dutt, A.K., Dunham, J.R. (1980). *Accident Research Manual*. University of North Carolina Highway Safety Research Center, Chapel Hill.
- [3] Dantas, A., Fowler, M., Koorey, G. (2007). Effect of road network bendiness on traffic crash occurrence. In *IPENZ Transportation Group Technical Conference*, Tauranga.
- [4] De Ceunynck, T., De Pauw, E., Daniels, S., Brijs, T., Hermans, E., Wets, G. (2012). *Identification of factors contributing to the occurrence of crashes at high-risk locations*. Policy Research Centre for Traffic Safety, zpráva RA-MOW-2011-027. Policy Research Centre for Traffic Safety, Diepenbeek.
- [5] Elvik, R. (2004). To what extent can theory account for the findings of road safety evaluation studies? *Accident Analysis and Prevention*, roč. 36, s. 841-849.
- [6] Elvik, R. (2011). Assessing causality in multivariate accident models. *Accident Analysis and Prevention*, roč. 43, s. 253-264.
- [7] Ferjenčík, J. (2000). *Úvod do metodologie psychologického výzkumu: Jak zkoumat lidskou duši*. Portál, Praha.
- [8] Gatti, G., Polidori, C., Galvez, I., Mallschützke, K., Jorna, R., van de Leur, M., Dietze, M., Ebersbach, D., Lippold, C., Schlag, B., Weller, G., Wyczynski, A., Iman, F., Aydin, C. (2007). *Safety Handbook for Secondary Roads*. Projekt RIPCORDER-ISEREST, zpráva D13.
- [9] Hauer, E. (2004). Statistical Road Safety Modeling. *Transportation Research Record*, č. 1897, s. 81-87.
- [10] Hendl, J. (2006). *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. 2. vydání. Portál, Praha.
- [11] Kubička, L. (1963). *Metody výzkumu*. In L. Fischer. Dětská psychiatrie pro mediky a lékaře. Státní zdravotnické nakladatelství, Praha.
- [12] Meloun, M., Militký, J. (2004). *Statistická analýza experimentálních dat*. 2. vydání. Academia, Praha.
- [13] Persaud, B.N. (2001). *Statistical Methods in Highway Safety Analysis: A Synthesis of Highway Practice*. NCHRP Synthesis 295. Transportation Research Board, Washington, USA.
- [14] Punch, K. (2008). *Základy kvantitativního šetření*. Portál, Praha.
- [15] Reurings, M., Janssen, T., Eenink, R., Elvik, R., Cardoso, J., Stefan, C. (2005). *Accident prediction models and road safety impact assessment: a state-of-the-art*. Projekt RIPCORDER-ISEREST, zpráva D2.1.
- [16] OECD (1997). *Road Safety Principles and Models: Review of Descriptive, Predictive, Risk and Accident Consequence Models*. OECD, Paříž.
- [17] Řeháková, B. (2000). Nebojte se logistické regrese. *Sociologický časopis*, roč. 36, č. 4, s. 475-492.

[18] Vieira Gomes, S., Geedipally, S.R., Lord, D. (2012). Estimating the safety performance of urban intersections in Lisbon, Portugal. *Safety Science*, roč. 50, s. 1732-1739.

[19] Hauer, E. *The art of regression modeling in road safety*. University of Toronto, 2014.

5.2 Seznam publikací, které předcházely metodice

Seznam obsahuje předcházející výstupy projektu IDEKO, seřazené podle času své publikace.

[20] Pokorný, P., Striegler, R. (2011). Identifikace nehodových lokalit. *Dopravní inženýrství*, roč. 6, č. 2, s. 8-11.

[21] Striegler, R., Pokorný, P. (2012). IDEKO – Nový přístup k identifikaci nehodových lokalit v ČR. In *XIII. Dopravně - inženýrské dny*, Hustopeče.

[22] Šenk, P., Pokorný, P., Ambros, J. (2012). Využití predikčního nehodovostního modelu při identifikaci kritických úseků na silniční síti. In *20. Silniční konference*, Plzeň, s. 112-116.

[23] Ambros, J., Pokorný, P. (2012). Identifikace nehodových lokalit – projekt IDEKO. In *Bezpečnost a plynulost v cestnej premávke*, Nový Smokovec.

[24] Ambros, J. (2012). Jak měřit bezpečnost? – 1. část. *Silniční obzor*, roč. 73, č. 3, s. 78-82.

[25] Ambros, J. (2012). Jak měřit bezpečnost? – 2. část. *Silniční obzor*, roč. 73, č. 4, s. 103-105.

[26] CDV (2012). *Identifikace kritických míst na pozemních komunikacích v extravilánu*. CDV, Brno.

[27] Šenk, P., Ambros, J., Pokorný, P., Striegler, R. (2012). Use of Accident Prediction Models in Identifying Hazardous Road Locations. *Transactions on Transport Sciences*, roč. 5, č. 4, s. 223-232.

[28] CDV (2013). *Řešení kritických míst na pozemních komunikacích v extravilánu*. CDV, Brno.

[29] Valentová, V. Difference between the expected numbers of accidents for various types of intersection arrangements in Prague. In *Young Researcher Seminar*, Lyon.

[30] Valentová, V. (2013). Accident Prediction Models for Three and Four Leg Priority Intersections in Prague. In *Modern Safety Technologies in Transportation*, Zlatá Idka.

[31] Valentová, V., Ambros, J., Janoška, Z. (2013). A Comparative Analysis of Identification of Hazardous Locations in Regional Rural Road Network. In *Road Safety and Simulation, International Conference RSS2013*, Řím.

[32] Pokorný, P., Ambros, J., Janoška, Z., Kyselý, M. (2013). Bezpečnost pozemních komunikací sítě TEN-T. *Silniční obzor*, roč. 74, č. 11, s. 10-16.

[33] Valentová, V., Ambros, J., Janoška, Z., Striegler, R. (2013). Predikční modelování a jeho výhody oproti stávající metodě identifikace nehodových lokalit. *Silniční obzor*, roč. 74, č. 11, s. 290-293.

5.3 Výstupy z originální práce

Seznam obsahuje výstupy, kterých bylo v roce 2014 dosaženo v souvislosti s aktivitami úzce souvisejícími s tématem metodiky. Přehled je platný k prosinci 2014.

[34] Valentová, V., Ambros, J., Janoška, Z. (2014). A comparative analysis of identification of hazardous locations in regional rural road network. *Advances in Transportation Studies*, č. 34, s. 57-66.

[35] Pokorný, P., Ambros, J. (2014). Identifikace rizikových míst na silnicích I. třídy v Jihomoravském kraji – praktická aplikace empirické bayesovské metody. *Silniční obzor*, roč. 75, č. 11 (v tisku).

[36] Ambros, J., Janoška, Z. (2015). Safety evaluation of Czech roundabouts (přijato k prezentaci na *94th Annual Meeting of the Transportation Research Board*, Washington).

[37] Ambros, J., Valentová, V., Janoška, Z. (2015). Investigation of difference between network screening results based on multivariate and simple crash prediction models (přijato k prezentaci na *94th Annual Meeting of the Transportation Research Board*, Washington).

Název **Multifaktorová analýza dopravní nehodovosti – metodika provádění**
Poskytovatel dotace: Ministerstvo vnitra ČR, Program bezpečnostního výzkumu České republiky v letech 2010-2015 (BV II/2 - VS)
Zhotovitel: **Centrum dopravního výzkumu, v. v. i.** (IČ: 44994575)
Autoři: Ing. Jiří Ambros, Ing. Veronika Valentová, Ing. Radim Striegler, Ing. Pavel Havránek, Ing. Eva Simonová, Ing. Martin Lipl, Ing. Lucie Vyskočilová, Ing. Petr Pokorný, Ing. Jan Novák, Mgr. Martin Kyselý, Mgr. Zbyněk Janoška, Ing. Ondřej Gogolín, Ing. Jindřich Frič, Ph.D.
Metodika schválena: Ministerstvem dopravy, č. j. 113/2014-710-VV/1
Oponenti: Doc. Ing. Josef Kocourek, Ph.D., ČVUT v Praze
Ing. Vlasta Michková, ŘSD
Vydalo: Centrum dopravního výzkumu, v. v. i. (IČ: 44994575)
Náklad: 200 ks
Tisk: Lubomír Červinka (IČ: 40981541)
Distribuce: Centrum dopravního výzkumu, v. v. i., Líšeňská 33a, 636 00 Brno
www.cdv.cz, cdv@cdv.cz, tel.: 541 641 711

© CDV, 2014, 1. vydání
ISBN 978-80-88074-01-4



www.cdv.cz