Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Studijní program: Kvantitativní metody v ekonomice

Studijní obor: Statistika a ekonometrie

ANALÝZA KONTINGENČNÍCH TABULEK V DOTAZNÍKOVÉM ŠETŘENÍ ŘIDIČŮ

Bakalářská práce

Autorka bakalářské práce: Barbora Velacková

Vedoucí bakalářské práce: Ing. Zdeněk Šulc

Akademický rok 2015/2016

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracovala samostatně a že jsem uvedla všechny použité prameny a literaturu, ze kterých jsem čerpala.

V Praze dne 2. 5. 2016

…….
podpis

Poděkování

Na tomto místě bych ráda poděkovala panu Ing. Zdeňku Šulcovi za pomoc při zpracování této bakalářské práce, zejména za jeho ochotu, cenné rady a poskytnutý materiál. Také bych chtěla poděkovat výzkumné agentuře Data Collect s.r.o., která mi vstřícně poskytla data k této práci.

Abstrakt

Bakalářská práce se zabývá analýzou dat z dotazníkového šetření řidičů pomocí kontingenčních tabulek. Data byla získána od výzkumné agentury Data Collect s.r.o., která provedla výzkum řidičů v roce 2014. Cílem práce je analyzovat chování řidičů a jejich návyky, které mohou zvyšovat riziko dopravních nehod. Práce je rozdělena do dvou hlavních částí, v první jsou popsány metody analýzy kontingenčních tabulek a ve druhé je samotná analýza dat. Nejprve je zkoumáno chování svobodných a mladých řidičů, v druhé části analýzy je porovnáváno chování mužů a žen za volantem. Výpočty byly provedeny ve statistickém programu SPSS a v tabulkovém procesoru MS Excel, kde byly vytvořeny také grafy a tabulky.

Klíčová slova: řidiči, kontingenční tabulky, šance, koeficienty asociace, dotazníkové šetření, testování hypotéz

Abstract

The bachelor thesis deals with the contingency table analysis from questionnaire survey data of drivers. The data were obtained from the agency Data Collect s.r.o., which conducted the survey in 2014. The aim of the thesis is to analyse the behaviour of drivers and their habits, which could increase the risk of accidents. The thesis is divided into two main parts; in the first one, methods of contingency table analysis are described; in the second one, the presented analyses are applied to the survey data. Firstly, the behaviour of single and young drivers is analysed, then the differences between men and women drivers. Calculations were made using the software SPSS and MS Excel, in which all the graphs and tables were made.

Key words: drivers, contingency tables, odds, association coefficient, questionnaire survey, hypothesis testing

Obsah

Úvod ... 2
1 Analýza kontingenčních tabulek ... 3
1.1 Dvourozměrná kontingenční tabulka ... 3
1.1.1 Chí-kvadrát test o vzájemné nezávislosti ... 4
1.1.2 Test nezávislosti věrohodnostním poměrem ... 5
1.1.3 Test o nulovosti Spearmanova koeficientu pořadové korelace ... 6
1.2 Šance ... 7
1.3 Míry intenzity závislosti dvou proměnných ... 8
1.3.1 Nominální proměnné ... 8
1.3.2 Ordinální proměnné ... 11
1.4 Čtyřpolní tabulka ... 12
1.5 Vícerozměrná kontingenční tabulka ... 14
2 Data ... 17
2.1 Svobodní řidiči ... 17
2.1.1 Řízení pod vlivem alkoholu ... 19
2.1.2 Telefonování za volantem ... 23
2.1.3 Rozčilování se na ostatní řidiče ... 26
2.1.4 Strach za volantem ... 28
2.1.5 Shrnutí ... 30
2.2 Genderová problematika u řidičů ... 30
2.2.1 Strach z řízení - neřidiči ... 31
2.2.2 Řízení jako obživa ... 33
2.2.3 Řízení pod vlivem alkoholu ... 34
2.2.4 Telefonování a psaní SMS za volantem ... 35
2.2.5 Zpívání za volantem ... 36
2.2.6 Shrnutí ... 38
Závěr ... 39
Přehled literatury ... 41
Internetové zdroje ... 41
Seznam tabulek ... 42
Seznam obrázků ... 42

Úvod

V dnešní době jsou automobily běžnou součástí života každého z nás, ať už z pohledu řidičů nebo chodců, ovlivňují náš každodenní život. Proto má význam zaobírat se chováním řidičů, zkoumat jejich společné rysy a odlišnosti v jednání a přístupu k pravidlům a hledat rizika vedoucí k zvýšené nehodovosti. Právě takové zkoumání je předmětem této práce, která se zaměřuje na chování řidičů za volantem z jejich perspektivy, na rozdíl od statistik nehodovosti, které nabízí jiný pohled.

Práce je empirická, zaměřená na aplikování metod analýzy kontingenčních tabulek na data z dotazníkového šetření řidičů. Data jsem získala od výzkumné agentury Data Collect s.r.o., která šetření řidičů sama provedla v roce 2014 metodou cawibus, tedy prostřednictvím webového formuláře, kam mohlo otázky vkládat více zadavatelů. Tato data obsahují zejména kategoriální proměnné, které lze dobře analyzovat právě pomocí kontingenčních tabulek.

Cílem mé práce je analyzovat chování řidičů za volantem poněkud subjektivněji z pohledu řidičů, kteří se do výzkumu zapojili, a získané výsledky otestovat, a tím je zobecnit na celou populaci. Speciálně se chci zaměřit na analýzu svobodných řidičů a genderovou problematiku, přičemž je mým záměrem popsat zejména chování a návyky, které mohou zvyšovat riziko vzniku nehod na silnicích, a porovnat je s publikovanými statistikami a obecnými předpoklady.

První část práce je teoretická, v níž jsou uvedeny metody analýzy kontingenčních tabulek, které jsem následně použila při analýze dat. Praktická část je členěna na dvě kapitoly, první se zaobírá svobodnými a mladými řidiči a druhá se zaměřuje na genderovou problematiku. U mladých a svobodných řidičů se předpokládá, že budou mít méně zkušeností a podle statistik právě tato kategorie řidičů zaviní velké procento nehod na silnicích, proto se zaměřím na odlišnosti v jejich jednání za volantem oproti ostatním řidičům. Je zjišťováno, zda jsou méně zodpovědní, než starší a zkušenější řidiči a tedy zda více riskují například řízením pod vlivem alkoholu, rychlou jízdou nebo telefonováním za volantem.

V druhé kapitole je zkoumána odlišnost mezi muži a ženami ve vztahu k řízení vozidel. Je zkoumáno, zda více riskují za volantem muži nebo ženy, kdo řídí častěji a jaké jsou příčiny neřízení vozidla. Rizikovými faktory bude zejména řízení pod vlivem alkoholu a telefonování a psaní zpráv za volantem.

Veškeré výpočty jsem prováděla ručně v tabulkovém procesoru Microsoft Excel a s pomocí statistického programu SPSS verze 20, kde jsem ruční výpočty především kontrolovala.

1 Analýza kontingenčních tabulek

1.1 Dvourozměrná kontingenční tabulka

Vhodným způsobem pro zkoumání souvislostí mezi dvěma kategoriálními proměnnými je kontingenční tabulka (tabulka 1.1). Jedná se o dvourozměrnou tabulku četností, kde kategorie proměnné X jsou uvedeny v řádcích x_i, kde $i = 1, 2 \ldots, r$ a kategorie proměnné Y jsou zapsané ve sloupcích y_j, kde $j = 1, 2 \ldots, s$. Uvnitř tabulky jsou hodnoty n_{ij}, pro $i = 1, 2 \ldots, r$ a $j = 1, 2 \ldots, s$, které představují sdružené absolutní četnosti udávající počet výskytů kombinace i-té kategorie proměnné X a j-té kategorie proměnné Y.

Četnosti v kontingenční tabulce mohou být tří typů – sdružené, které již byly představeny, marginální a podmíněné. Marginální četnosti jsou součty řádků sdružených absolutních četností, které se značí n_{i+}, nebo součty sloupců sdružených absolutních četností, jež se značí n_{+j}. Každé marginální rozdělení přísluší jedné proměnné z kontingenční tabulky.

Tabulka 1.1 Kontingenční tabulka – symbolika

	y_1	y_2	…	y_s	n_{i+}
x_1	n_{11}	n_{12}	…	n_{1s}	n_{1+}
x_2	n_{21}	n_{22}	…	n_{2s}	n_{2+}
…	…	…	…	…	…
x_r	n_{r1}	n_{r2}	…	n_{rs}	n_{r+}
n_{+j}	n_{+1}	n_{+2}	…	n_{+s}	n

Zdroj: (*Pecáková*, 2011).

Kontingenční tabulka může obsahovat kromě absolutních četností také četnosti relativní. Sdružené relativní četnosti

$$p_{ij} = n_{ij} / n \tag{1.1}$$

vyjadřují podíl výskytu kombinace daných kategorií v souboru a celkového počtu. Řádkové marginální relativní četnosti

$$p_{i+} = n_{i+} / n \tag{1.2}$$

zobrazují podíl kategorií proměnné X v souboru a sloupcové marginální relativní četnosti

$$p_{+j} = n_{+j} / n \tag{1.3}$$

vyjadřují podíl kategorií proměnné Y.

V mnoha kontingenčních tabulkách je jedna proměnná vysvětlovaná (tedy závislá) a druhá vysvětlující (nezávisle) proměnná. V takovém případě je vhodné uvést zvlášť rozdělení

vysvětlované proměnné na každé úrovni vysvětlující proměnné, které se nazývá podmíněné rozdělení. Souvislosti mezi proměnnými pak lze pozorovat pomocí podmíněných relativních četností v řádcích tabulky

$$p_{j/i} = p_{ij} / p_{i+} = n_{ij} / n_{i+} \tag{1.4}$$

nebo pomocí podmíněných relativních četností ve sloupcích tabulky

$$p_{i/j} = p_{ij} / p_{+j} = n_{ij} / n_{+j} \,. \tag{1.5}$$

Určení závislosti či nezávislosti proměnných v kontingenční tabulce je podstatnou součástí analýzy dat. K tomu lze použít různé testy nezávislosti, které jsou popsány níže, pomocí nichž je možné zamítnout nebo nezamítnout testovanou hypotézu. Statistická hypotéza je určitý předpoklad, který jsme vyslovili o určité charakteristice nebo tvaru rozdělení v populaci, jak je uvedeno v (Hindls a kol., 2006).

Postup testování spočívá v tom, že se nejprve stanoví testovaná (nulová) hypotéza H_0 a v rozporu s ní alternativní hypotéza H_1. Ta může být oboustranná, která obecně tvrdí, že charakteristika v nulové hypotéze, často se jedná o parametr určitého rozdělení, se nerovná zadané hodnotě, nebo může být jednostranná. V takovém případě se rozlišuje levostranná alternativa, kde je charakteristika menší, než za platnosti nulové hypotézy, a nebo pravostranná alternativa, kde je charakteristika větší.

Cílem testování nulové hypotézy je určit, zda lze či nelze na určité hladině významnosti zamítnout nulovou hypotézu ve prospěch alternativy. Hladina významnosti je pravděpodobnost chyby prvního druhu, což je pravděpodobnost, že je nesprávně zamítnuta nulová hypotéza, která ve skutečnosti platí, a značí se α. Obvykle se stanovuje $\alpha = 0{,}05$ nebo $\alpha = 0{,}01$. Pokud se nesprávně nezamítne nulová hypotéza, přestože neplatí, jedná se o chybu druhého druhu, která se značí β, jak je uvedeno v (Řezanková, 2007).

1.1.1 Chí-kvadrát test o vzájemné nezávislosti

Cílem chí-kvadrát testu nezávislosti je zjištění, zda jsou na sobě dvě kategoriální proměnné nějakým způsobem závislé, či nikoliv. Vhodný je pak především u proměnných nominálních. Testuje se nulová hypotéza H_0*: proměnné jsou nezávislé* oproti alternativní hypotéze H_1*: proměnné jsou závislé*. Jako testové kritérium zde slouží Pearsonova statistika chí-kvadrát daná vztahem

$$X^2 = \sum_i^s \sum_j^r \frac{(n_{ij} - m_{ij})^2}{m_{ij}}, \tag{1.6}$$

kde m_{ij} jsou teoretické (očekávané) četnosti výskytu v *i*-té a *j*-té kategorii při výběru o rozsahu *n*, které se spočítají jako

$$m_{ij} = \frac{n_{i+} n_{+j}}{n}. \tag{1.7}$$

Chí-kvadrát test je založen na porovnávání empirických četností s teoretickými, přičemž za platnosti nulové hypotézy o nezávislosti proměnných by se empirické četnosti měly co nejvíce shodovat s teoretickými. Z toho vyplývá i opačná situace, čím větší je rozdíl mezi teoretickými a empirickými četnostmi, tím vyšší je hodnota testového kritéria X^2 a tím větší je šance na zamítnutí nulové hypotézy, jak uvádí Agresti (2007).

Aby však byla zachována přesnost aproximace chí-kvadrát rozdělením, teoretické četnosti by v jednotlivých polích kontingenční tabulky neměly klesnout pod hodnotu dvě a alespoň v 80 % polí by neměly klesnout pod hodnotu pět. Pokud jsou tyto podmínky splněny, nemělo by docházet ke zkreslení výsledků.

Kritickou hodnotou je $(1-\alpha)\%$ kvantil chí-kvadrát rozdělení s $(r-1)(s-1)$ stupni volnosti, tedy $\chi^2_{1-\alpha}(r-1)(s-1)$. Pokud je hodnota testového kritéria vyšší, než je příslušná kritická hodnota, lze na α % hladině významnosti zamítnout nulovou hypotézu ve prospěch hypotézy alternativní, a tím prokázat závislost mezi proměnnými.

1.1.2 Test nezávislosti věrohodnostním poměrem

Další možností ověřování nezávislosti dvou kategoriálních proměnných je použití testu věrohodnostním poměrem, který se v případě multinomického rozdělní vypočítá jako

$$G^2 = 2\sum_i^r \sum_j^s n_{ij} \ln \frac{n_{ij}}{m_{ij}} \tag{1.8}$$

a konverguje také k chí-kvadrát rozdělení s $(r-1)(s-1)$ stupni volnosti. U tohoto testu jsou taktéž předpoklady určitého minimálního počtu polí v tabulce, aby byla zajištěna přijatelná aproximace. Teoretické četnosti by měly být vyšší než pět alespoň v 80 % polí v kontingenční tabulce a zároveň by každá teoretická četnost měla nabývat hodnoty vyšší než jedna.

Pokud nejsou v důsledku malého rozsahu výběru dodržena tato kritéria, nelze považovat aproximaci těchto statistik za vyhovující. Řešením může být například sloučení některých

kategorií, což zmenší rozměry kontingenční tabulky a může mít dopad na samotný výsledek testu. Dalším řešením může být použití Fisherova exaktního testu, který je vysvětlen například v (Agresti, 2007), přičemž je výpočetně náročnější.

Pearsonova statistika X^2 a věrohodnostní poměr G^2 jsou sice odlišnými testovými statistikami, ale mnoho vlastností mají společných a většinou poskytují shodné závěry. Pokud je nulová hypotéza správná a rozsah výběru je velký, tyto dvě statistiky mají shodné chí-kvadrát rozdělení a nabývají téměř shodných hodnot, jak uvádí Agresti (2007).

1.1.3 Test o nulovosti Spearmanova koeficientu pořadové korelace

K určení lineární nezávislosti dvou ordinálních proměnných slouží Spearmanův koeficient pořadové korelace, který vychází z určení průměrného pořadí jednotlivých hodnot obou veličin. Pořadová čísla jsou přiřazována jednotkám souboru, které je vhodné uspořádat podle stupně sledovaného znaku. Spearmanův koeficient se pak vypočítá podle vzorce

$$r_S = \frac{\sum_i \sum_j R'(i) R'(j) n_{ij}}{\frac{1}{12}\sqrt{\left(n^3 - \sum_i n_{i+}^3\right)\left(n^3 - \sum_j n_{+j}^3\right)}}, \tag{1.9}$$

kde $i = 1, 2, \ldots, r$ a $j = 1, 2, \ldots, s$, a nabývá hodnot od -1 do 1. Krajních hodnot nabývá pouze ve specifickém případě, kdy se přesně shodují jednotlivá pořadí obou veličin, a to ve čtvercové tabulce. V čitateli se nachází $R(i)$ a $R(j)$, což jsou průměrná pořadí jednotlivých hodnot veličin a Y, která se vypočítají jako

$$R(i) = \sum_{k<i} n_{k+} + \frac{n_{i+} + 1}{2}, \tag{1.10}$$

$$R(j) = \sum_{k<j} n_{+k} + \frac{n_{+j} + 1}{2}. \tag{1.11}$$

Protože však hodnota Spearmanova koeficientu není ovlivněna libovolnou monotónní transformací pořadí, je možno zjednodušit zápis koeficientu tak, aby platilo

$$\sum_i R'(i) n_{i+} = \sum_j R'(j) n_{+j} = 0,$$

přičemž hodnoty $R'(i)$ a $R'(j)$ jsou získány odečtením celkového průměrného pořadí $\frac{n+1}{2}$.

Následně vzniknou vzorce

$$R'(i) = \sum_{k<i} n_{k+} + \frac{n_{i+} - n}{2}, \tag{1.12}$$

$$R'(j) = \sum_{k<j} n_{+k} + \frac{n_{+j} - n}{2}, \tag{1.13}$$

které již lze dosadit do čitatele Spearmanova koeficientu pořadové korelace (1.9). Tento koeficient je vlastně Pearsonovým korelačním koeficientem průměrných pořadí. Otestování nulovosti Spearmanova koeficientu pořadové korelace se provádí pomocí testového kritéria

$$t = r_S \sqrt{\frac{n-2}{1-r_S^2}}, \tag{1.14}$$

které má při platnosti nulové hypotézy, jež říká, že je tento koeficient nulový, Studentovo t rozdělení s (n – 2) stupni volnosti, jak uvádí Řezanková (2007).

1.2 Šance

Šance a poměr šancí jsou důležitými mírami asociace mezi proměnnými. Pomocí absolutních nebo relativních četností lze vyjádřit šanci úspěchu, případně neúspěchu, tedy šanci, že veličina Y nabude j-té a ne j'-té hodnoty. Ta se určí srovnáním podmíněných relativních četností této veličiny

$$\frac{p_{j/i}}{p_{j'/i}} = \frac{p_{ij}}{p_{ij'}} = \frac{n_{ij}}{n_{ij'}} \tag{1.15}$$

za podmínky, že četnosti ve jmenovateli jsou vždy větší než nula. Srovnají-li se šance ve dvou řádcích tabulky, vyjde poměr šancí, který se pro i-tou a i'-tou hodnotu veličiny X vypočítá jako

$$\Theta_{ij} = \frac{p_{j/i} / p_{j'/i}}{p_{j/i'} / p_{j'/i'}} = \frac{p_{ij} p_{i'j'}}{p_{ij'} p_{i'j}} = \frac{n_{ij} n_{i'j'}}{n_{ij'} n_{i'j}}, \tag{1.16}$$

jak uvádí Pecáková (2011). Poměr šancí nabývá nezáporných hodnot a interpretuje se přímo pomocí zlomku. V případě nezávislosti jsou šance pro obě proměnné shodné a poměr šancí je tedy roven 1. Tato hodnota je určitým mezníkem mezi dvěma typy asociace mezi kategoriemi proměnných. Pokud je poměr šancí větší než 1, pak je šance na úspěch vyšší u prvního řádku tabulky, respektive toho řádku, jehož šance byla brána do poměru šancí jako první, a naopak. Dále platí, že čím dále od jedné se hodnota poměru šancí nachází, tím je silnější závislost mezi

proměnnými. Pokud se v kontingenční tabulce zamění řádky se sloupci, nemá to na poměr šancí žádný vliv, stejně tak v případě násobení řádků či sloupců konstantou.

Poměr šancí má silně zešikmené rozdělení, jelikož je nezávislost dána hodnotou 1, od níž poměr šancí může klesat pouze k nule, zatímco na opačnou stranu může růst nade všechny meze. Z tohoto důvodu se místo šance může použít přirozený logaritmus šance, zvaný logit

$$\ln \frac{p_{j/i}}{p_{j'/i}} = \ln \frac{p_{ij}}{p_{ij'}}, \tag{1.17}$$

jež je ekvivalentní mírou k šanci. Poměr logaritmů šancí může nabývat libovolné hodnoty a jeho rozdělení je symetrické okolo nuly, která značí nezávislost, jak uvádí Agresti (2007).

Zamění-li se řádky tabulky mezi sebou, má to vliv pouze na změnu znaménka, nikoli na hodnotu poměru logaritmu šancí, neboť platí

$$\ln \Theta = -\ln(1/\Theta). \tag{1.18}$$

1.3 Míry intenzity závislosti dvou proměnných

Je-li zjištěna závislost mezi dvěma proměnnými, je vhodné určit také intenzitu této závislosti. K měření síly závislosti slouží různé koeficienty popsané v následujících kapitolách, pro jejichž interpretaci je vhodné, aby nabývaly hodnot od 0 do 1 u nominálních proměnných, případně od -1 do 1 u proměnných ordinálních, u nichž lze zjistit i směr závislosti. Hodnoty blízké nule pak znamenají spíše nezávislost či malou závislost veličin, zatímco blíží-li se hodnoty k jedné, značí to silnou závislost veličin.

Pro volbu vhodné míry intenzity závislosti je důležitý zejména typ proměnné, tedy zda je proměnná nominální (jmenná) či ordinální (pořadová). Dále je podstatný tvar kontingenční tabulky, jež může být čtvercová, u které k sobě lze jednoznačně přiřadit dvojice proměnných, nebo může být obdélníková, u které lze jednoznačně přiřadit pouze hodnoty jedné proměnné hodnotám druhé. Konkrétně pro situaci, kdy r > s, je možné jednoznačně přiřadit hodnoty *Y* hodnotám *X*, ale nelze to provést opačně. Záleží také na tom, zda rozlišujeme směr závislosti či nikoliv.

1.3.1 Nominální proměnné

Mírami intenzity závislosti nominálních proměnných jsou koeficienty kontingence. Jako první jsou zde uvedeny takové, které se zakládají na Pearsonově statistice chí-kvadrát

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij} - n_{i+}n_{+j}/n\right)^2}{n_{i+}n_{+j}/n} = n\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(p_{ij} - p_{i+}p_{+j}\right)^2}{p_{i+}p_{+j}}, \tag{1.19}$$

jejíž hodnota roste spolu s rozsahem souboru, a proto není sama o sobě vhodnou mírou intenzity závislosti. Koeficienty založené na této statistice nabývají v případě nezávislosti hodnoty 0, jejich maxima se však liší. Jedná se o symetrické míry, což znamená, že vyjadřují vzájemné závislosti dvou proměnných.

Mezi nejznámější patří Pearsonův koeficient kontingence

$$C_P = \sqrt{\frac{X^2}{X^2 + n}}, \tag{1.20}$$

který nabývá hodnot z intervalu $\left\langle 0, \sqrt{h/(h+1)} \right\rangle$, kde h = min (r – 1, s – 1), přičemž uvedené maximum je menší než jedna. Pokud je Pearsonův koeficient kontingence vydělen svým maximem, lze dosáhnout snazší interpretace, neboť hodnoty výsledného C_P^*, kde

$$C_P^* = \frac{C_P}{\max(C_P)}, \tag{1.21}$$

leží v intervalu od 0 do 1. Ve stejném intervalu leží také hodnota Cramérova koeficientu *V*, která již ve jmenovateli obsahuje maximální možnou hodnotu statistiky chí-kvadrát

$$V = \sqrt{X^2 / nh}. \tag{1.22}$$

Obecně je doporučeno používat Cramérovo *V* spíše v tabulkách obsahujících alespoň jednu alternativní proměnnou (proměnná nabývající pouze hodnot 0 nebo 1), zatímco přepočtený Pearsonův koeficient kontingence v ostatních případech.

Další mírou intenzity závislosti dvou proměnných je *koeficient průměrné čtvercové kontingence* (*phi*), který lze vypočítat pomocí vzorce

$$\varphi = \sqrt{X^2 / n} \tag{1.23}$$

a jehož maximální hodnotou je $\sqrt{h}$. Pokud se jedná o čtyřpolní tabulku (2x2), hodnota koeficientu *phi* závislých proměnných bude rovna jedné, v jiných případech může být i větší než jedna a narůstá spolu s rozměry tabulky.

Následující míry intenzity závislosti se již nezakládají na Pearsonově statistice, ale porovnávají variabilitu marginálního rozdělení jedné proměnné a průměrnou variabilitu

podmíněných rozdělení této veličiny. Tyto míry asociace již nejsou symetrické, lze tedy určit směr závislosti proměnných.

Goodman Kruskalovo tau

$$\tau_{yx} = \frac{\sum_i \sum_j p_{ij}^2 / p_{i+} - \sum_j p_{+j}^2}{1 - \sum_j p_{+j}^2} \tag{1.24}$$

vyjadřuje míru závislosti *Y* na *X*, přičemž je použita Giniho průměrná odchylka jakožto marginální variabilita. Giniho průměrná odchylka udává „podíl počtu odlišných dvojic jednotek na počtu všech možných dvojic. Jedná se vlastně o průměrnou diferenci, která v případě soustředění všech jednotek do jedné kategorie nabývá hodnoty 0, v případě rovnoměrného zastoupení všech *K* kategorií v souboru pak hodnoty (*K* - 1)/*K*.“[1]

Úpravou vzorce se dá zjistit také závislost *X* na *Y*. Pokud by byla místo Giniho průměrné odchylky použita entropie, což je míra neurčitosti, která zároveň představuje množství informace, které je potřebné k jejímu odstranění, je možné pro výpočet míry závislosti použít koeficient neurčitosti neboli informační koeficient

$$U_{yx} = \frac{\sum_i \sum_j p_{ij} \ln\left(p_{i+} p_{+j} / p_{ij}\right)}{\sum_j p_{+j} \ln p_{+j}} . \tag{1.25}$$

Tyto dva koeficienty často nabývají nízkých hodnot v důsledku rostoucích rozměrů kontingenční tabulky, což lze upravit odmocněním koeficientů, čímž se přiblíží hodnotám koeficientů kontingence.

Další mírou intenzity je Goodman Kruskalovo lambda

$$\lambda_{yx} = \frac{\sum_i \max(p_{ij}) - \max(p_{+j})}{1 - \max(p_{+j})} , \tag{1.26}$$

jehož konstrukce se opírá o modus veličiny. Pokud je *Y* nezávislé na *X*, jsou všechny největší četnosti výskytu ve stejném sloupci tabulky.

[1] PECÁKOVÁ, Iva. (2011). *Statistika v terénních průzkumech*. 2. vyd. Praha: Professional Publishing. s. 80.

1.3.2 Ordinální proměnné

V kapitole týkající se nominálních proměnných byla statistická závislost označována jako kontingence, v případě ordinálních proměnných se jedná o *korelaci*. Odlišnost spočívá v tom, že kategorie ordinálních proměnných mohou být jednoznačně seřazeny podle stupně sledovaného znaku, což vede k jinému principu počítání. Lze rozlišovat korelaci *pozitivní*, kdy nízkým hodnotám jedné proměnné odpovídají nízké hodnoty druhé proměnné, nebo naopak korelaci *negativní*, kdy nízkým hodnotám jedné proměnné odpovídají vysoké hodnoty proměnné druhé, jak uvádí Pecáková (2011) nebo Řezanková (2007).

Před uvedením samotných měr závislosti je nutno nejprve zmínit několik pojmů. Ze souboru o rozsahu n je možné sestavit $n(n-1)/2$ párů jednotek, které mohou být trojího typu. Za prvé může být tato dvojice shodná v hodnotách obou veličin a počet těchto dvojic se vypočítá pomocí sdružených četností jako

$$T_{yx} = \frac{1}{2}\sum_i \sum_j n_{ij}\left(n_{ij} - 1\right). \tag{1.27}$$

Za druhé se dvojice proměnných může shodovat pouze v hodnotě jedné veličiny. Jsou-li dvojice shodné pouze v hodnotách X, jejich počet se určí jako

$$T_x = \sum_i \sum_j n_{ij} \sum_{h=i} \sum_{k>j} n_{hk}\ , \tag{1.28}$$

zatímco shodují-li se pouze v hodnotách Y, použije se následující vzorec

$$T_y = \sum_i \sum_j n_{ij} \sum_{h>i} \sum_{k=j} n_{hk}\ . \tag{1.29}$$

Ve třetím případě mohou být hodnoty obou veličin odlišné. Rozlišují se páry konkordantní, které jsou charakteristické vyššími hodnotami obou proměnných u jedné jednotky, než u druhé jednotky, a páry diskordantní, u nichž je hodnota jedné proměnné vyšší a druhé proměnné nižší u jedné jednotky, než u jednotky druhé. Počet konkordantů se určí jako

$$C = \sum_i \sum_j n_{ij} \sum_{h>i} \sum_{k>j} n_{hk} \tag{1.30}$$

a počet diskordantů následovně

$$D = \sum_i \sum_j n_{ij} \sum_{h>i} \sum_{k<j} n_{hk}\ . \tag{1.31}$$

Následující koeficienty asociace vycházejí z výše uvedených vzorců pro počty párů jednotek.

Mezi symetrické míry patří Goodmanův a Kruskalův koeficient asociace gama

$$\Gamma = \frac{C - D}{C + D}, \tag{1.32}$$

který se zakládá na počtu konkordantních a diskordantních párů. V absolutní hodnotě nabývá hodnot od 0 do 1, přičemž 0 znamená nezávislost a hodnota 1 úplnou závislost, která nastává, pokud jsou v kontingenční tabulce nenulové hodnoty pouze na diagonále.

Další symetrickou mírou intenzity závislosti ordinálních proměnných je Kendallovo τ_B (tau-b), neboli Kendallův koeficient pořadové korelace, viz (Pecáková, 2011), jenž je dán vztahem

$$\tau_B = \frac{C - D}{\sqrt{(C + D + T_x)(C + D + T_y)}} = \frac{2 \cdot (C - D)}{\sqrt{\left(n^2 - \sum_i n_{i+}^2\right)\left(n^2 - \sum_j n_{+j}^2\right)}} \tag{1.33}$$

a nabývá hodnot od -1 do 1. Mezi symetrické míry se řadí i Kendallovo τ_C (tau-c), které se vypočítá jako

$$\tau_C = \frac{C - D}{C_{\max}} = \frac{2 \cdot (C - D)}{n^2 (m - 1) / m}, \tag{1.34}$$

přičemž $m = \min(r, s)$. Kendallovo τ_C nabývá taktéž hodnot od -1 do 1. (Pecáková, 2011, Řezanková, 2007).

Mezi asymetrické míry intenzity závislosti ordinálních proměnných patří Somersův koeficient d, který lze vypočítat jako

$$d_{yx} = \frac{C - D}{C + D + T_y} = \frac{2 \cdot (C - D)}{n^2 - \sum_i n_{i+}^2}. \tag{1.35}$$

Tento koeficient nabývá hodnot od -1 do 1 a má i svou symetrickou variantu, která je harmonickým průměrem obou asymetrických variant koeficientu. Geometrickým průměrem těchto asymetrických variant koeficientu d je Kendallovo τ_B, které bylo původně představeno v (Řezanková, 2007).

1.4 Čtyřpolní tabulka

Pokud se v kontingenční tabulce nachází dvě alternativní proměnné, jedná se o čtyřpolní tabulku, pro niž lze mnohdy vzorce měr intenzity závislosti zjednodušit. Statistiku chí-kvadrát (1.6) je možno pro čtyřpolní tabulku zjednodušit jako

$$X^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1+}n_{+1}n_{2+}n_{+2}}. \tag{1.36}$$

Dále v tomto specifickém případě platí, že je Cramérův koeficient shodný s koeficientem *phi* a je vždy vyšší, než Pearsonův koeficient, jehož maximum je 0,707, jak uvádí Pecáková (2011).

Jelikož proměnné ve čtyřpolní tabulce nabývají pouze dvou hodnot, je možné určit směr závislosti, stejně jako u ordinálních proměnných. Na tuto tabulku lze tedy aplikovat některé koeficienty, které jsou uvedeny v předchozí kapitole o ordinálních proměnných. Například koeficient asociace gama (1.25) lze zapsat jako

$$Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}} \tag{1.37}$$

a je pro čtyřpolní tabulku známý jako Yuleův koeficient Q. Dále pak Kendallovo τ_B (1.26) lze zjednodušit na tvar

$$\tau_B = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1+}n_{+1}n_{2+}n_{+2}}} \tag{1.38}$$

a symetrická varianta Somersova koficientu *d* do podoby

$$d = \frac{2(n_{11}n_{22} - n_{12}n_{21})}{n_{1+}n_{2+} + n_{+1}n_{+2}}, \tag{1.39}$$

přičemž tyto dva koeficienty nabývají ve čtyřpolní tabulce téměř stejných hodnot. Pokud by tabulka byla symetrická podle nějaké diagonály, tak by nesymetrické Somersovy koeficienty byly shodné.

Pro čtyřpolní tabulku lze zjednodušit také poměr šancí (1.16) následovně

$$\Theta = \frac{n_{11}n_{22}}{n_{12}n_{21}}, \tag{1.40}$$

přičemž jej nelze určit v případě nulových četností ve jmenovateli. Tento problém lze však odstranit přičtením malé kladné konstanty ke všem četnostem. Poměr šancí nabývá, na rozdíl od ostatních měr intenzity asociace, hodnoty jedna v případě nezávislosti a nezáporné hodnoty v případě závislosti. Také je nesymetrickou mírou, takže pokud se změní směr závislosti, poměr šancí nabude převrácené hodnoty. Tyto problémy lze eliminovat zlogaritmováním šancí, jak uvádí Pecáková (2011).

1.5 Vícerozměrná kontingenční tabulka

Dvourozměrné kontingenční tabulky mohou být v určitých případech nedostatečné. Souvislost dvou proměnných může komplikovat nějaká další veličina, jejímž vlivem se zkoumané dvě veličiny pouze zdají být závislé, nebo naopak mezi nimi nějaká souvislost je, ale další veličina tuto souvislost zakrývá. K rozpoznání těchto složitějších vztahů slouží vícerozměrné kontingenční tabulky. Zde bude popsána tabulka trojrozměrná.

Trojrozměrná kontingenční tabulka je na rozdíl od dvourozměrné tříděna navíc ještě podle třetí kategoriální proměnné *Z*. Z toho vyplývá, že kromě *r* řádků a *s* sloupců obsahuje tabulka navíc *v* vrstev. Dvourozměrně lze takovou tabulku zobrazit například následovně (tabulka 1.2).

Tabulka 1.2 Vícerozměrná kontingenční tabulka

		y_1	y_2	$\dots$	y_s	n_{i+k}
z_1	x_1	n_{111}	n_{121}	$\dots$	n_{1s1}	n_{1+1}
	x_2	n_{211}	n_{221}	$\dots$	n_{2s1}	n_{2+1}
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	x_r	n_{r11}	n_{r21}	$\dots$	n_{rs1}	n_{r+1}
	n_{+j1}	n_{+11}	n_{+21}	$\dots$	n_{+s1}	n_{++1}
z_2	x_1	n_{112}	n_{122}	$\dots$	n_{1s2}	n_{1+2}
	x_2	n_{212}	n_{222}	$\dots$	n_{2s2}	n_{2+2}
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	x_r	n_{r12}	n_{r22}	$\dots$	n_{rs2}	n_{r+2}
	n_{+j2}	n_{+12}	n_{+22}	$\dots$	n_{+s2}	n_{++2}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
z_v	x_1	n_{11v}	n_{12v}	$\dots$	n_{1sv}	n_{1+v}
	x_2	n_{21v}	n_{22v}	$\dots$	n_{2sv}	n_{2+v}
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	x_r	n_{r1v}	n_{r2v}	$\dots$	n_{rsv}	n_{r+v}
	n_{+jk}	n_{+1v}	n_{+2v}	$\dots$	n_{+sv}	n_{++v}
	n_{+j+}	n_{+1+}	n_{+2+}	$\dots$	n_{+s+}	n

Zdroj: (*Pecáková*, 2011)

Pokud se v trojrozměrné kontingenční tabulce sloučí řádky, sloupce nebo vrstvy, vznikne marginální dvourozměrné rozdělení, ve kterém je ignorována třetí proměnná. Lze zobrazit také podmíněná dvourozměrná rozdělení dvou veličin, která jsou podmíněná veličinou třetí.

Vztahy mezi třemi veličinami jsou o něco komplikovanější, než závislosti mezi dvěma proměnnými. Mohou mezi nimi navzájem existovat až čtyři různé typy vztahů, které budou následně uvedeny.

Všechny tři proměnné mohou být vzájemně nezávislé. Taková situace nastává, pokud dvě veličiny, například *X* a *Y*, zůstávají nezávislé i poté, co jsou roztříděné podle třetí veličiny *Z*, takže poměr šancí je jednotkový. V takovém případě jsou podmíněná rozdělení *X* a *Y* shodná s marginálním rozdělením. To stejné platí pro podmíněná a marginální rozdělení ostatních dvojic veličin *X, Z* a *Y, Z*.

Asociace dvou proměnných je nezávislá na třetí veličině. Pokud třetí veličina, například *Z*, neovlivňuje ani jednu veličinu samostatně a ani jejich souvislost, lze ji v analýze vynechat. Taková situace se pozná podle shodných podmíněných rozdělení *X* a *Y* s rozdělením marginálním, ale poměr šancí nabývá jiné hodnoty, než jedna. Veličiny *X* a *Y* jsou tedy vzájemně závislé.

Dvojice veličin může být podmíněně nezávislá, což znamená, že vlivem třetí veličiny se proměnné sice zdají být závislé, ale nejsou. Z podmíněných rozdělení *X* a *Y* vyplývá pro všechny hodnoty *Z* nezávislost této dvojice, ale z marginálního rozdělení plyne jejich zdánlivá závislost, kterou lze vypozorovat z marginálního poměru šancí, který je odlišný od jedné. Oba podmíněné poměry šancí jsou ale rovny přibližně jedné, což značí nezávislost. Tato situace vzniká, pokud jsou obě veličiny, tedy *X* a *Y*, závislé na *Z*. Chceme-li tedy zkoumat vztah prvních dvou veličin, je nutno vliv proměnné *Z* eliminovat.

Všechny tři proměnné mohou být po dvojicích vzájemně závislé. Pokud se podmíněná rozdělení navzájem liší, ale podmíněné poměry šancí jsou ve dvojicích shodné, jsou veličiny navzájem závislé, jak uvádí Pecáková (2011).

Zjištění nezávislosti mezi veličinami je v případě vícerozměrné kontingenční tabulky vítané, jelikož se může vliv nezávislé proměnné eliminovat, a tím poté zmenšit rozměry tabulky. Pokud se například podmíněná rozdělení shodují, lze namísto nich použít rozdělení marginální, což zjednoduší celou analýzu.

Jak již bylo řečeno, vztahy mezi proměnnými ve vícerozměrné kontingenční tabulce se mohou jevit jinak, než jaké doopravdy jsou. Proto by použití měr asociace pro dvourozměrné kontingenční tabulky mohlo být zavádějící. Z tohoto důvodu existují parciální míry asociace, pomocí nichž je možné eliminovat vliv jedné nebo více veličin, a tím docílit výsledku, který by lépe odpovídal skutečným vztahům mezi proměnnými.

V soustavě čtvercových kontingenčních tabulek je možno pro vyjádření parciální asociace použít parciální korelační koeficienty. Pokud se jedná například o asociaci dvou alternativních veličin X a Y, přičemž je potřeba eliminovat vliv veličiny Z, lze vypočítat parciální korelační koeficient prvního řádu jako

$$r_{yx.z} = \frac{r_{yx} - r_{yz} r_{xz}}{\sqrt{\left(1 - r_{yz}^2\right)\left(1 - r_{xz}^2\right)}}, \tag{1.41}$$

přičemž obecně písmeno uvedené v dolním indexu za tečkou značí veličinu, jejíž vliv je eliminován. Parciální korelační koeficient je funkcí koeficientů řádu nula, tedy párových korelačních koeficientů. Samotný Pearsonův koeficient korelace se pro dvě alternativní proměnné vypočte jako

$$r_{yx} = \frac{n_{11} n_{22} - n_{12} n_{21}}{\sqrt{n_{1+} n_{+1} n_{2+} n_{+2}}}. \tag{1.42}$$

Stejným způsobem by bylo možno určit parciální korelační koeficienty vyššího řádu, kde by se eliminoval vliv více veličin, a jednalo by se o funkci korelačních koeficientů vyšších řádů, jak uvádí Pecáková (2011).

Roztřídí-li se čtyřpolní tabulka do v vrstev dle hodnot další proměnné, lze určit stupeň parciální asociace stanovením MH společného poměru šancí jako

$$\Theta_{MH} = \frac{\sum_{k=1}^{v} \frac{n_{11k} n_{22k}}{n_k}}{\sum_{k=1}^{v} \frac{n_{12k} n_{21k}}{n_k}}, \tag{1.43}$$

přičemž MH označuje Mantelovu-Haenszelovu statistiku. Tu lze použít pouze v případě, že se charakter závislosti v jednotlivých vrstvách moc neliší, jinak by mohlo dojít ke zkreslení výsledků.

2 Data

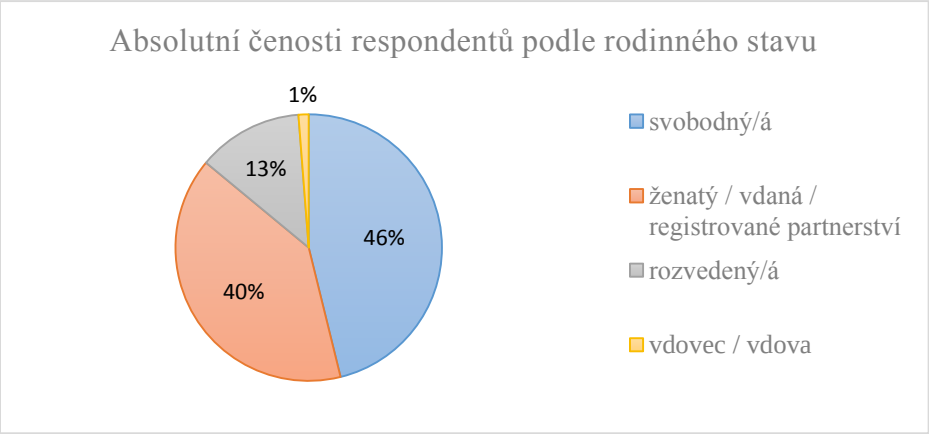
Data k této práci byla získána od výzkumné agentury Data Collect s.r.o., která provedla v roce 2014 internetový průzkum řidičů. Společnost použila metodu CAWIBUS, neboli kombinaci CAWI (Computer Assisted Web Interviewing), tedy dotazování pomocí webového formuláře, a omnibus. „Omnibus je název pro výzkumné dotazování, při kterém si do dotazníku vloží otázky více zadavatelů. Tím se dělí o náklady, což je velmi výhodné, i pokud máte pouze několik otázek a samostatný výzkum by byl nákladný. Dotazník tak obsahuje více témat a každý zadavatel dostane výsledky za svou část a demografické charakteristiky respondentů.“[2]

Jednalo se o kvótní výběr na online panelu dané společnosti, konkrétně se jednalo o kvóty na pohlaví, věk a region vůči aktivní online populaci ČR. Tento výzkum tedy není reprezentativní vůči obecné populaci, což ale platí obecně při použití online metodiky sběru dat. Sama společnost si pak z výzkumu odnesla zejména základní představu o názorech řidičů a tato studie byla východiskem pro další fáze výzkumného projektu řidičů.

Databáze obsahuje odpovědi od 650 respondentů a je v ní 72 proměnných, které jsou ve většině případů kategoriální.

2.1 Svobodní řidiči

První analýza této práce analyzuje chování svobodných řidičů. Tato skupina zahrnuje řidiče, kteří jsou z větší části ve věku 18 – 30let, a proto by mohla přibližně zastupovat kategorii mladých a méně zkušených řidičů. Skupinu tvoří i starší, ale svobodní řidiči. U svobodných respondentů se předpokládá menší zodpovědnost při řízení v porovnání s řidiči, kteří jsou v manželství a mají potomky, proto jsou zde uvedeni v jedné kategorii spolu s mladými. Pro získání základní představy o výběru jsou zde nejprve uvedeny některé grafy.

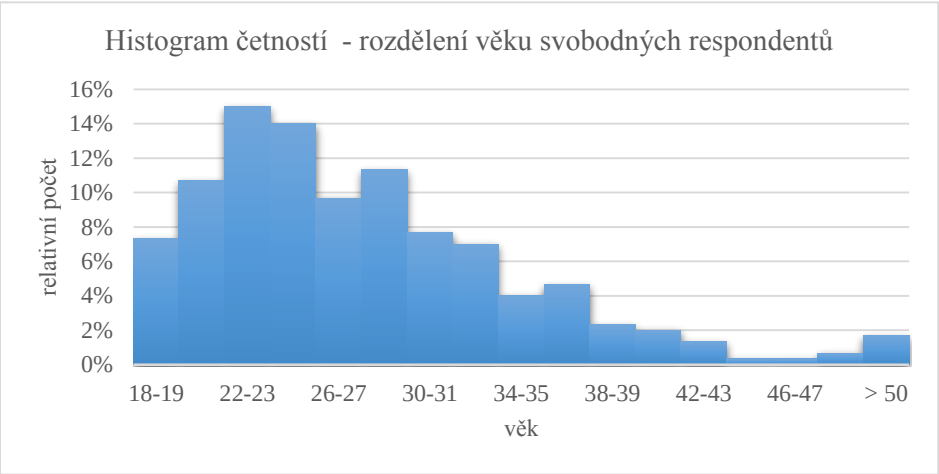


Obrázek 1.1 Výsečový graf – rodinný stav

[2] Data Collect s.r.o., omnibus [online]

Na obrázku 1.1 lze pozorovat výsečový graf absolutních četnosti respondentů rozčleněných podle rodinného stavu. Svobodných řidičů je v tomto výběru nejvíce, jedná se tedy o modální kategorii a z celkového počtu 650 respondentů tvoři tato skupina téměř polovinu.

Na obrázku 1.2 lze vidět rozdělení věku zvolené kategorie svobodných respondentů. Rozdělení je zešikmené vlevo, proto tíhne k nižším hodnotám a je zde názorně ukázáno, že většina svobodných řidičů ve vzorku je spíše nižšího věku. Proto byla pro zastoupení mladých a méně zodpovědných řidičů zvolena právě tato kategorie.



Obrázek 1.2 Histogram – věk svobodných respondentů

Jak již bylo naznačeno v úvodu, kategorie mladých řidičů má nezanedbatelný podíl na dopravní nehodovosti. Velmi mladí řidiči (do 25 let) zavinili v ČR v období 2007 – 2010 přibližně 21 % všech dopravních nehod a zapříčinili až 25 % úmrtí při dopravních nehodách, jak uvádí Český statistický úřad[3]. Řidiči od 25 do 35 let pak zavinili v témže období dalších 30 % kolizí a způsobili 27 % úmrtí vlivem dopravních nehod.

V datech, jež jsou zdrojem této analýzy, nejsou uvedeny informace o nehodovosti respondentů, a proto bude zkoumána spíše zodpovědnost řidičů vyplývající z jejich chování za volantem a přístupu k dodržování pravidel. Jedním z porušovaných pravidel je řízení automobilu pod vlivem alkoholu.

Měření hladiny alkoholu u řidičů probíhá z větší části až u dopravních nehod. Jak uvádí Český statistický úřad[4], za rok 2013 byla zjištěna stopa alkoholu nebo drog u téměř 5 tisíc

[3] KAMENICKÝ, J. (2014). Mladí a nezkušení řidiči jako rizikový faktor nehodovosti. *Dopravní nehodovost a její důsledky v ČR v dlouhodobém pohledu* [online], ČSÚ.
[4] KAMENICKÝ, J. (2014). Vliv alkoholu na dopravní nehodovost. *Dopravní nehodovost a její důsledky v ČR v dlouhodobém pohledu* [online], ČSÚ.

viníků dopravních nehod. V porovnání s celkovým počtem nehod, u kterých bylo možné testovat přítomnost alkoholu, tak bylo v roce 2013 v ČR zaviněno 6,4 % nehod.

V následující podkapitole jsou zkoumány odpovědi respondentů týkající se právě řízení pod vlivem alkoholu, které nemohou být evidovány v žádných statistikách přesně a zakládají se na předpokladu pravdivě uváděných informací v důsledku anonymity dotazníků.

2.1.1 Řízení pod vlivem alkoholu

Hlavním záměrem této podkapitoly je zkoumat, zda se vyskytuje nějaká závislost mezi řízením automobilu pod vlivem alkoholu a tím, zda respondent spadá do zvolené kategorie svobodných, či nikoliv.

Pro účely analýzy byla proměnná rodinný stav upravena. Byla ponechána skupina „svobodný/á“, ale ostatní kategorie byly sloučeny do jedné s názvem „ostatní“. Sestavenou kontingenční tabulku zobrazující absolutní četnosti výskytu kombinací jednotlivých kategorií lze vidět v tabulce 2.1.

Tabulka 2.1 Kontingenční tabulka – rodinný stav, alkohol

		Rodinný stav		
		Svobodný/á	Ostatní	Celkem
Řídil/a jste někdy vůz pod vlivem alkoholu?	Ano, vícekrát	15	30	45
	Ano, jednou	46	52	98
	Ne	235	265	500
	Celkem	296	347	643

Jako první je proveden chí-kvadrát test o nezávislosti dvou proměnných, přičemž je testována nulová hypotéza H_0: *nezávislost proměnných* oproti alternativní hypotéze H_1: *závislost proměnných*. Poté jsou podle (1.7) vypočteny teoretické četnosti v kontingenční tabulce, přičemž je potvrzeno splnění předpokladů pro provedení chí-kvadrát testu.

Dále je podle (1.6) vypočteno testové kritérium $X^2 = 3{,}14$, které je porovnáno s kvantilem $\chi^2_{0,95}(2) = 5{,}99$. Jelikož je hodnota testového kritéria nižší, než kritická hodnota, testové kritérium spadá do oboru přijetí a nelze tedy na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti proměnných. Podle (1.7) je vypočteno testové kritérium pro test věrohodnostním poměrem $G^2 = 3{,}22$, ale ani to není vyšší, než výše uvedený kvantil. Závislost mezi řízením pod vlivem alkoholu a rodinným stavem nebyla prokázána.

Nelze tedy říci, že by svobodní respondenti řídili pod vlivem alkoholu více, než ostatní respondenti. To může být způsobeno například tím, že jsou v této kategorii mladší a méně zkušení řidiči, kteří ještě nedůvěřují svým řidičským dovednostem natolik, aby si troufali sedat

za volant po požití alkoholu. Další příčinou by mohla být méně častá pravdivost odpovědí v dotazníku u mladších respondentů, což je ale naneštěstí nezjistitelné, a tak tato možnost nebude brána v potaz. Nelze tedy vyloučit možnost, že řízení pod vlivem alkoholu opravdu není závislé na věku či rodinném stavu řidičů a někteří řidiči napříč všemi generacemi tak ohrožují zdraví ostatních.

Dala by se však předpokládat nějaká závislost mezi řízením v podnapilém stavu a pohlavím svobodného řidiče. V tabulce 2.2 se nachází absolutní četnosti svobodných mužů a žen a jejich odpovědi na otázku řízení pod vlivem alkoholu.

Tabulka 2.2 Kontingenční tabulka – pohlaví, alkohol

		Pohlaví		
		Muž	Žena	Celkem
Řídil/a jste někdy vůz pod vlivem alkoholu?	Ano, vícekrát	13	2	15
	Ano, jednou	34	12	46
	Ne	118	117	235
	Celkem	165	131	296

Stejně jako v předchozím příkladu byl proveden chí-kvadrát test o nezávislosti proměnných. Testové kritérium vypočtené podle (1.6) nabývá hodnoty $X^2 = 14{,}88$, což je jednoznačně vyšší, než kritická hodnota $\chi^2_{0,95}(2) = 5{,}99$. Na 5% hladině významnosti lze zamítnout nulovou hypotézu o nezávislosti proměnných ve prospěch alternativní hypotézy. Tyto dvě proměnné na sobě závislé jsou.

Jaká je mezi nimi závislost je možné zjistit výpočtem měr intenzity závislosti pro dvě nominální proměnné. V tabulce 2.3 se nachází vypočtené symetrické koeficienty podle (1.20) – (1.23).

Tabulka 2.3 Výpočty – symetrické koeficienty – pohlaví, alkohol

C_P	C_P^*	V	ϕ
0,219	0,309	0,224	0,224

Hodnoty Cramérova V a koeficientu průměrné čtvercové kontingence jsou stejné, neboť se v kontingenční tabulce vyskytuje proměnná Pohlaví, která má jen dvě kategorie, a tudíž je hodnota $h = min(r - 1)(s - 1) = 1$. V takovém případě se od sebe vzorce (1.22) a (1.23) neliší. Z tabulky 2.3 lze pozorovat, že hodnoty měr intenzity závislosti jsou vcelku nízké, takže závislost mezi řízením vozu pod vlivem alkoholu a pohlavím je docela malá.

Dále jsou vypočteny nesymetrické míry intenzity závislosti, které určují pouze jednostrannou závislost řízení pod vlivem alkoholu na pohlaví. Zjišťování opačné závislosti by nemělo věcnou interpretaci. Závislou proměnnou je tedy řízení pod vlivem alkoholu a

nezávislou proměnnou pohlaví. Podle vzorců (1.24) – (1.26) byly vypočteny následující koeficienty, které jsou v tabulce 2.4.

Tabulka 2.4 Výpočty - nesymetrické koeficienty – pohlaví, alkohol

τ_{yx}	U_{yx}	λ_{yx}
0,035	0,044	0,000

Z tabulky 2.4 lze vidět, že všechny nesymetrické koeficienty nabývají velmi nízkých hodnot. Goodman Kruskalovo tau a koeficient neurčitosti jsou téměř shodné a vyplývá z nich, že pohlaví má na řízení pod vlivem alkoholu jen nepatrný vliv.

Lambda je rovna nule, neboť se všechny maximální četnosti nacházejí ve stejném řádku kontingenční tabulky. Modální je tedy kategorie těch řidičů, kteří nikdy neřídili pod vlivem alkoholu, což lze pozorovat již z tabulky absolutních četností (tabulka 2.2). Většina respondentů tedy nikdy neřídila pod vlivem alkoholu, což je jistě pozitivní zjištění, avšak je nutno zmínit, že téměř 20,6 % svobodných řidičů nejméně jednou v podnapilém stavu za volant usedlo.

Názornějšího srovnání řízení pod vlivem alkoholu podle pohlaví je možno docílit pomocí šancí. Šance, že bude pod vlivem alkoholu řídit muž, ať už jednou, nebo vícekrát, je přibližně 2:5 oproti situaci, kdy muž nebude řídit podnapilý. U žen je pak tato šance více než 1:10. Z poměru těchto šancí pak vyplývá, že pokud bude řidič řídit pod vlivem alkoholu, bude to s více než 3krát vyšší šancí muž.

Dalším předmětem zájmu je zjistit, zda řidiči, kteří požili alkohol před usednutím za volant, riskují i jinými způsoby. V úvahu přichází odpověď na dotazníkovou otázku: Zkoušel/a jste někdy s vozem, jaká je jeho maximální rychlost, kterou dokáže jet? Opět jsou zvlášť zkoumány odpovědi svobodných respondentů a ostatních řidičů. Pro přehlednější zobrazení byly proměnné zjednodušeny do alternativní podoby a k analýze byla použita trojrozměrná kontingenční tabulka (tabulka 2.5).

Tabulka 2.5 Trojrozměrná kontingenční tabulka – alkohol, rychlost, rodinný stav

Rodinný stav			Alkohol		Celkem
			Ano	Ne	
Svobodný/á	Rychlost	Ano	39	80	119
		Ne	21	153	174
	Celkem		60	233	293
Ostatní	Rychlost	Ano	49	82	131
		Ne	33	183	216
	Celkem		82	265	347
Celkem			142	498	640

Po vypočtení podmíněného rozdělení a marginálního rozdělení, jež je očištěno o vliv veličiny *Z*, v tomto případě rodinný stav, lze vidět, že se tato rozdělení téměř shodují. Dále je třeba určit, zda jsou na sobě všechny tři veličiny nezávislé nebo zda existuje závislost mezi proměnnými *X* a *Y*. Podle (1.42) jsou vypočteny párové korelační koeficienty pro každou dvojici proměnných. Je-li brána veličina značící rychlost jako *x*, alkohol jako *y* a rodinný stav jako *z*, bylo spočítáno r_{xy} = 0,251, r_{yz} = -0,038 a r_{xz} = 0,029.

Na základě párových korelačních koeficientů byl dopočítán podle (1.41) parciální korelační koeficient eliminující veličinu *z* jako $r_{xy.z}$ = 0,252. Lze konstatovat, že bez vlivu rodinného stavu je závislost mezi alkoholem za volantem a testováním rychlosti nepatrně vyšší.

Následně byly vypočteny poměry šancí v jednotlivých vrstvách, tedy zvlášť pro svobodné řidiče Θ_1 = 3,552 a pro ostatní řidiče Θ_2 = 3,314. Poměry šancí jsou velmi podobné, proto je možné dopočítat podle (1.43) MH společný poměr šancí Θ_{MH} = 3,405, který říká, že šance řízení pod vlivem alkoholu pro ty řidiče, kteří zkoušeli maximální rychlost vozu, je více než 3:1 vzhledem k situaci, kdy pod vlivem alkoholu neřídí. V tabulce 2.6 jsou uvedeny absolutní četnosti pouze pro tyto dvě kategorie, tedy bez rodinného stavu.

Tabulka 2.6 Čtyřpolní tabulka – alkohol, rychlost

		Alkohol		
		Ano	Ne	Celkem
Rychlost	Ano	88	162	250
	Ne	54	336	390
Celkem		142	498	640

Nejprve je proveden chí-kvadrát test o nezávislosti proměnných. Hodnota testového kritéria je X^2 = 40,24 a je následně porovnávána s kvantilem $\chi^2_{0,95}(1)$ = 3,84. Jak lze vidět, hodnota testového kritéria je mnohonásobně vyšší, než příslušný kvantil, a proto lze na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti proměnných. Stejného výsledku by bylo docíleno provedením testu věrohodnostním poměrem, kde je testové kritérium vypočtené podle (1.7) G^2 = 39,44 taktéž vyšší, než příslušný kvantil.

Určitá závislost mezi řízením pod vlivem alkoholu a testováním rychlé jízdy existuje, a proto jsou zde dále vypočteny míry intenzity závislosti, které přiblíží, o jak velkou závislost se jedná. Tyto koeficienty jsou vypočteny podle vzorců (1.20) – (1.23) a jejich hodnoty jsou uvedeny v tabulce 2.7. Cramérovo *V* nabývá stejné hodnoty jako koeficient průměrné čtvercové kontingence (*phí*), což platí obecně pro čtyřpolní tabulku. Pearsonův koeficient je nepatrně

nižší, než tyto koeficienty, ale po přepočtení je nejlépe interpretovatelný. Na škále od nuly do jedné nabývá přibližně 0,344, což značí sice slabou závislost, ale nezanedbatelnou.

Tabulka 2.7 Výpočty – koeficienty pro čtyřpolní tabulku – alkohol, rychlost

C_p	C_p^*	V	ϕ
0,243	0,344	0,251	0,251

Dále byly podle (1.15) vypočteny šance, že respondenti řídili pod vlivem alkoholu, a to zvlášť pro ty, kteří testovali maximální rychlost vozu a ty, kteří ne. U řidičů, kteří zkoušeli maximální rychlost vozidla, je šance 1:2, že také někdy řídili pod vlivem alkoholu. Naproti tomu u řidičů, kteří nejvyšší rychlost vozidla nezkoušeli, je tato šance přibližně 1:6, tedy znatelně nižší. Poměr těchto šancí byl již vyjádřen výše.

2.1.2 Telefonování za volantem

V této podkapitole bylo u svobodných respondentů zkoumáno, zda na sobě nějakým způsobem závisí telefonování za volantem bez handsfree a úroveň řidičských schopností, kterou měli sami respondenti určit podle uvedené škály. Jedná se o ordinální veličiny, tudíž je k testování nezávislosti použitý Spearmanův koeficient pořadové korelace a jsou použity míry intenzity závislosti, které zohledňují pořadí kategorií proměnných. Absolutní četnosti odpovědí respondentů lze vidět v tabulce 2.8.

Telefonování za volantem je další rizikovou činností, která je sice zakázána, ale mnoho řidičů tak činí poměrně často. K takovému závěru lze dojít prostým chvilkovým pozorováním řidičů na silnicích a křižovatkách. Už z prvního pohledu na sloupcové marginální četnosti proměnné telefonování za volantem v tabulce 2.8 lze vidět, že více než polovina respondentů nejméně jednou telefonovala bez handsfree při řízení automobilu. Dále je v této kapitole zkoumáno, zda je nějaká závislost mezi telefonováním za volantem a sebevědomím řidičů, respektive jejich důvěrou ve své řidičské schopnosti.

Tabulka 2.8 Kontingenční tabulka– Jak byste charakterizoval/a své řidičské schopnosti, telefonování

	Jak často telefonujete bez handsfree, když řídíte?					
	Nikdy	Výjimečně	Občas	Často	Téměř pokaždé	Celkem
Špičkový/á řidič/ka	3	1	1	2	1	8
Velmi dobrý/á řidič/ka	23	27	19	2	0	71
Spíše dobrý/á řidič/ka	37	27	8	2	0	74
Průměrný/á řidič/ka	43	21	7	1	0	72
Spíše špatný/á řidič/ka	6	2	1	0	0	9
Celkem	112	78	36	7	1	234

Nejprve je vhodné určit, zda jsou na sobě tyto dvě ordinální veličiny závislé, což lze zjistit pomocí Spearmanova koeficientu korelace pořadí. Jako první jsou vypočtena průměrná pořadí

jednotlivých hodnot veličin podle (1.10) a (1.11), od nichž se následně odečte celkové průměrné pořadí. Tím se získala $R'(i)$ a $R'(j)$ podle (1.12) a (1.13), která už lze dosadit do čitatele vzorce pro Spearmanův koeficient (1.9). Tento koeficient vyšel r_S = -0,26.

Pomocí testu na nulovost Spearmanova koeficientu lze otestovat závislost mezi proměnnými. Testové kritérium je vypočtené podle (1.14) t = -4,140. V důsledku velkého rozsahu výběru lze Studentovo t rozdělení nahradit normovaným normálním rozdělením, a tudíž je zde porovnáváno testové kritérium t s kvantilem $u_{0,975}$ = 1,96. V absolutní hodnotě je testové kritérium vyšší, a proto lze na 5% hladině významnosti zamítnout nulovou hypotézu o nulovosti Spearmanova korelačního koeficientu. Proměnné jsou na sobě závislé.

Dále jsou vypočteny podle vzorců (1.27) – (1.31) počty dvojic shodných v obou hodnotách proměnných, počty dvojic shodných v hodnotách jedné proměnné a dále počty konkordantů a diskordantů, které jsou následně využity při výpočtu měr intenzity závislosti mezi proměnnými. Počty těchto různých dvojic lze vidět v tabulce 2.9.

Tabulka 2.9 Počty dvojic ordinálních proměnných – řidičské schopnosti, telefonování

T_{yx}	T_x	T_y	C	D
2 976	4 830	6 894	4 138	8 423

Pomocí hodnot z tabulky 2.9 je možno podle (1.32) – (1.35) dopočíst míry intenzity závislosti zvolených ordinálních proměnných. Vypočtené koeficienty jsou v tabulce 2.10.

Tabulka 2.10 Výpočty – koeficienty ordinálních proměnných – řidičské schopnosti, telefonování

Γ	τ_B	τ_C	d_{yx}	d_{xy}
-0,341	-0,233	-0,196	-0,220	-0,246

Všechny vypočtené koeficienty nabývají záporných hodnot, což vzhledem k uspořádání kategorií v kontingenční tabulce (tabulka 2.8) znamená, že ti svobodní respondenti, kteří se označují za spíše lepší řidiče, častěji telefonují bez handsfree, než ti, kteří se označují za průměrné či spíše horší řidiče. Jednou z příčin může být to, že svým schopnostem více důvěřují, a tak nabývají dojmu, že zvládnou řídit stejně dobře i při telefonování. Naproti tomu řidiči, kteří svým schopnostem tolik nedůvěřují, jsou za volantem opatrnější, a tedy si moc netroufají rozptylovat se telefonováním bez handsfree.

Dalším nebezpečným trendem, který se v posledních letech rozmáhá, je psaná SMS za volantem, které je ještě nebezpečnější, než telefonování za jízdy, a to v důsledku potřeby odtrhnout zrak od vozovky. Předmětem zkoumání tedy bude, zda za volantem píší SMS více svobodní nebo zda na rodinném stavu nezáleží. Vzhledem k tomu, že do kategorie svobodných

spadá mnoho mladých, kteří jsou v dnešní době mobily téměř posedlí, dala by se určitá míra závislosti předpokládat.

Dále je zkoumána závislost mezi psaním SMS a telefonováním za volantem, tedy zda je u těch řidičů, kteří někdy telefonovali za volantem, vyšší šance, že budou také psát zprávy. V následující trojrozměrné tabulce 2.11 jsou absolutní četnosti výskytu jednotlivých kategorií těchto tří proměnných.

Tabulka 2.11 Trojrozměrná kontingenční tabulka – rodinný stav, telefonování, SMS

			Psaní SMS za volantem		
			Ne	Ano	Celkem
Svobodný/á	Telefonování za volantem	Ne	100	12	112
		Ano	47	75	122
		Celkem	147	87	234
Ostatní	Telefonování za volantem	Ne	126	12	138
		Ano	86	56	142
		Celkem	212	68	280
Celkem			359	155	514

Porovnáním podmíněných rozdělení s marginálním bylo zjištěno, že jsou téměř shodné, ale podmíněné poměry šancí se nerovnají jedné, z čehož lze usuzovat na závislost dvou proměnných bez vlivu třetí proměnné. Dále jsou podle (1.42) vypočteny párové korelační koeficienty, z nichž bude patrné, jaké jsou vztahy mezi dvojicemi veličin. Jestliže je telefonování za volantem x, psaní SMS za volantem y a rodinný stav z, párové korelační koeficienty jsou r_{xy} = 0,436, r_{yz} = -0,140 a r_{xz} = -0,014. Pomocí těchto koeficientů byly podle (1.41) dopočítány parciální korelační koeficienty, které eliminují vliv třetí veličiny, jako $r_{xy.z}$ = 0,438, $r_{yz.x}$ = -0,149 a $r_{xz.y}$ = 0,053. Z těchto výsledků vyplývá středně silná pozitivní závislost mezi telefonováním a psaním zpráv za volantem a také slabá negativní závislost mezi rodinným stavem a psaním zpráv za volantem.

Pro názornější vyjádření vztahu mezi psaním SMS za volantem a rodinným stavem byly vypočteny šance podle (1.15) a následně poměr těchto šancí podle (1.16). Šance, že bude při řízení psát SMS svobodný člověk, je přibližně 3:5, v případě ostatních je tato šance o něco nižší, a to 1:3. Poměr těchto šancí je téměř 2:1 a říká, že pokud řidič píše SMS zprávy za volantem, bude to s téměř dvojnásobnou šancí svobodný člověk. To potvrzuje počáteční domněnku o rizikovějším chování svobodných řidičů, alespoň co se psaní SMS za volantem týče.

Následující výpočty se vztahují pouze k první části tabulky 2.11, která se týká pouze svobodných respondentů. Je tedy zkoumána závislost mezi telefonováním a psaním SMS za

volantem u svobodných řidičů, kde se již z výsledných hodnot korelačních koeficientů uvedených výše předpokládá pozitivní středně silní závislost.

Je-li proveden test nezávislosti proměnných podle (1.6) pomocí testového kritéria chí-kvadrát X^2 = 64,42 nebo podle (1.7) pomocí věrohodnostního poměru G^2 = 69,92, lze dojít ke stejnému závěru. Obě testová kritéria jsou mnohem větší, než kvantil $\chi^2_{0,95}(1)$ = 3,84, a tedy lze na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti mezi telefonováním a psaním SMS za volantem. V následující tabulce 2.12 jsou vypočteny míry intenzity závislosti těchto dvou proměnných.

Tabulka 2.12 Výpočty – koeficienty nominálních proměnných – telefonování, psaní SMS

C_p	$V = \phi$	$\tau_{yx} = \tau_{xy}$	U_{yx}	U_{xy}	λ_{yx}	λ_{xy}
0,465	0,525	0,275	0,542	0,508	0,473	0,322

Hodnoty vypočtených koeficientů potvrzují středně silnou závislost mezi těmito veličinami. Byly vypočteny i nesymetrické koeficienty pro oba směry závislosti a lze pozorovat, že závislost telefonování na psaní SMS je o něco větší, než závislost opačná. To lze vysvětlit tím, že psaní zpráv za volantem je o něco rizikovější, tudíž pokud svobodní řidiči již riskovali psaním SMS, pak pravděpodobně nebudou mít větší problém s o něco méně rizikovým telefonováním za volantem. Zatímco ti svobodní řidiči, kteří někdy za volantem telefonovali, ještě mohou mít nějaké zábrany k rizikovějšímu psaní SMS za volantem.

Pro názornější vyjádření těchto vztahů byly vypočteny šance podle (1.15). Pokud svobodný řidič za volantem nepíše SMS, pak je šance téměř 1:2, že při řízení telefonuje. Jestliže ale svobodný řidič píše zprávy za volantem, pak je šance více než 6:1, že při řízení také telefonuje. Z poměru šancí vypočteného podle (1.16) pak vyplývá, že pokud svobodný řidič za volantem telefonuje, pak bude s více než 13krát vyšší šancí také psát SMS. Lze tedy říci, že je riziko telefonování a psaní SMS za volantem u svobodných řidičů poměrně vysoké.

2.1.3 Rozčilování se na ostatní řidiče

Za volantem se řidiči mohou chovat různě. Někteří jsou nervózní a roztěkaně sledují vozovku, jiní jsou uvolnění a trpěliví, a někteří jsou nevrlí a často se rozčilují. Právě na posledně zmíněné je zaměřena tato podkapitola. Nezvládání vlastních emocí za volantem může být také rizikovým faktorem, který může ohrozit bezpečí na silnicích, proto je zde sledováno, jak si v rozčilování za volantem stojí zvolená kategorie svobodných a ve většině případů mladších řidičů v porovnání s ostatními řidiči.

Jako první je tedy zkoumáno, zda je nějaká závislost mezi rozčilováním se u svobodných řidičů a ostatních řidičů. Absolutní četnosti odpovědí rozčleněné do dvou kategorií podle rodinného stavu lze vidět v tabulce 2.13. Podle empirických četností se zdálo, že bude porušen předpoklad pro teoretické četnosti, ale ty byly nižší, než 5 právě ve 20 % případů, což je únosné.

Tabulka 2.13 Kontingenční tabulka – rozčilování za volantem, rodinný stav

		Jak často se při řízení rozčilujete na ostatní řidiče?					Celkem
		Téměř pokaždé	Často	Občas	Výjimečně	Nikdy	
Rodinný stav	Svobodný/á	3	49	96	76	10	234
	Ostatní	4	35	128	110	3	280
	Celkem	7	84	224	186	13	514

Jelikož se jedná o kombinaci nominální a ordinální proměnné, byl použit chí-kvadrát test o nezávislosti dvou proměnných. Testové kritérium podle (1.6) je rovno $X^2 = 13{,}02$ a nabývá tedy vyšší hodnoty než kvantil $\chi^2_{0,95}(4) = 9{,}488$. Na 5% hladině významnosti lze zamítnout nulovou hypotézu o nezávislosti těchto dvou proměnných.

Tabulka 2.14 Výpočty – koeficienty nominálních proměnných rozčilování se a rodinný stav

C_P	C_P^*	V	ϕ	τ_{xy}	U_{xy}	λ_{xy}
0,153	0,216	0,155	0,155	0,006	0,011	0,000

V tabulce 2.14 jsou vypočteny symetrické i nesymetrické míry intenzity závislosti těchto dvou proměnných podle (1.20 – 1.26). Ze symetrických koeficientů lze pozorovat, že je vzájemná závislost mezi proměnnými vcelku malá. Uvedené nesymetrické koeficienty určují pouze jednostranný vztah, v tomto případě závislost četnosti rozčilování se za volantem na rodinném stavu respondenta.

Hodnoty nesymetrických koeficientů, Goodman Kruskalova tau a koeficientu neurčitosti, jsou téměř nulové a lambda nabývá nuly v důsledku stejné modální kategorie pro svobodné řidiče i ostatní. Tyto výsledky je možné celkově shrnout tak, že míra rozčilování se za volantem je téměř nezávislá na tom, zda jsou řidiči svobodní, či nikoliv. A tedy nelze říci, že by se jedni rozčilovali za volantem o moc častěji, než druzí.

U rozčilujících se řidičů je dále zkoumáno, zda umí být také nápomocni ostatním řidičům a upozorňují je například na nerozsvícená světla. Tentokrát jsou zkoumány odpovědi pouze u svobodných respondentů. Otázkou je, jestli a jak závisí rozčilování se řidičů a upozorňování ostatních řidičů na nerozsvícená světla. Absolutní četnosti kombinací odpovědí je možné vidět v tabulce 2.15.

Tabulka 2.15 Kontingenční tabulka – rozčilování se a upozorňování ostatních řidičů

		Jak často se při řízení rozčilujete na ostatní řidiče?					
		Nikdy	Výjimečně	Občas	Často	Téměř pokaždé	Celkem
Jak často upozorňujete ostatní řidiče, že např. nemají rozsvícená světla?	Nikdy	2	4	5	1	0	12
	Výjimečně	5	31	26	12	1	75
	Občas	2	33	49	19	1	104
	Často	1	5	10	15	1	32
	Téměř pokaždé	0	3	6	2	0	11
	Celkem	10	76	96	49	3	234

Pro analýzu této tabulky lze použít metody určené pro dvě ordinální proměnné. Stejně jako v příkladu o telefonování za volantem bez handsfree je nutné nejprve spočítat Spearmanův koeficient korelace pořadí podle (1.9). Spearmanův koeficient vychází $r_S = 0{,}24$ což značí slabou pozitivní závislost. Je vypočteno testové kritérium podle (1.14) $t = 3{,}75$, které je v důsledku velkého rozsahu výběru porovnáno s kvantilem $u_{0,975} = 1{,}96$, z čehož vyplývá, že lze na 5% hladině významnosti zamítnout nulovou hypotézu o nulovosti Spearmanova koeficientu, tedy o nezávislosti proměnných. Čím častěji se svobodný řidič za volantem rozčiluje na ostatní řidiče, tím častěji je také upozorňuje, že nemají rozsvícená světla.

Tabulka 2.16 Výpočty – koeficienty – rozčilování se a upozorňování na nerozsvícená světla

Γ	τ_B	τ_C
0,307	0,211	0,190

V tabulce 2.16 jsou vypočteny míry intenzity závislosti těchto dvou ordinálních veličin podle (1.32 – 1.34). Jelikož koeficient gama poněkud nadhodnocuje, lze říci, že je míra závislosti mezi proměnnými vcelku malá. Somersovy nesymetrické koeficienty d zde nejsou uvedeny, stejně jako symetrická podoba tohoto koeficientu, jelikož v tomto případě nabývají stejné hodnoty, jako Kendallovo τ_B, konkrétně 0,211. Závěrem lze říci, že pozitivní směr závislosti těchto dvou veličin značí fakt, že si rozčilující se řidiči více všímají chování ostatních řidičů na silnicích, a to jak ve špatném slova smyslu, tak v tom dobrém.

2.1.4 Strach za volantem

Předmětem zkoumání v této podkapitole je nejprve zjištění, zda je nějaká závislost mezi strachem z řízení automobilu u svobodných a u ostatních řidičů, přičemž absolutní četnosti odpovědí respondentů jsou uvedeny v tabulce 2.17, a dále, jestli se svobodní řidiči, kteří se řídit bojí, se svým strachem vypořádávají například vyhýbáním se větším a komplikovanějším křižovatkám, či nikoliv.

Tabulka 2.17 Kontingenční tabulka – strach za volantem, rodinný stav

		Jak často při řízení máte strach?				Celkem
		Nikdy	Výjimečně	Občas	Často	
Rodinný stav	Svobodný/á	27	98	72	37	234
	Ostatní	49	120	77	34	280
	Celkem	76	218	149	71	514

Vzhledem k porušení předpokladu teoretických četností pro provedení chí-kvadrát testu byly sloučeny kategorie „často“ a „téměř pokaždé“ do jedné. Testové kritérium vypočtené podle (1.6) $X^2 = 4{,}805$ bylo nižší, než hodnota kvantilu $\chi^2_{0,95}(3) = 7{,}815$, není možné na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti těchto dvou proměnných. Nelze tedy říci, že by se podíl svobodných řidičů, kteří se někdy při řízení bojí, výrazně odlišoval od ostatních řidičů.

Přesto je ale možné zaměřit se nyní na svobodné řidiče a zkoumat, zda se ti, kteří se častěji při řízení bojí, také častěji vyhýbají městu a komplikovanějším křižovatkám, či nikoliv. Absolutní četnosti kombinací odpovědí svobodných respondentů zobrazuje tabulka 2.18.

Tabulka 2.18 Kontingenční tabulka – strach z řízení, vyhýbání se městu

		Jak často se při řízení vyhýbáte větším městům nebo komplikovaným křižovatkám?					Celkem
		Nikdy	Výjimečně	Občas	Často	Téměř pokaždé	
Jak často při řízení máte strach?	Nikdy	14	11	2	0	0	27
	Výjimečně	29	41	22	6	0	98
	Občas	7	24	29	10	2	72
	Často	6	9	14	7	0	36
	Téměř pokaždé	0	0	0	1	0	1
	Celkem	56	85	67	24	2	234

Jelikož se jedná o dvě ordinální proměnné, je zde podle (1.9) vypočten Spearmanův koeficient pořadové korelace $r_S = 0{,}387$, který ukazuje na pozitivní závislost mezi těmito dvěma proměnnými. Již z tabulky 2.18 lze vidět, že se absolutní četnosti odpovědí vyskytují ve větší míře na diagonále, tedy s rostoucím strachem při řízení vzrůstá také četnost vyhýbání se větším městům a komplikovaným křižovatkám. Tato závislost je prokázána na 5% hladině významnosti provedením testu o nulovosti Spearmanova koeficientu pořadové korelace, kdy testové kritérium vypočtené podle (1.14) $t = 6{,}384$ je prokazatelně vyšší, než kvantil $u_{0,975} = 1{,}96$. Závislost proměnných lze také potvrdit výpočtem měr intenzity závislosti pro ordinální proměnné, které lze vidět v tabulce 2.19.

Tabulka 2.19 Výpočty – koeficienty ordinálních proměnných – strach z řízení, vyhýbání se městu

Γ	τ_B	τ_C	d_{yx}	d_{xy}
0,465	0,335	0,295	0,341	0,329

Na základě analýz provedených v této podkapitole vyplývá, že se svobodní řidiči, kteří se více bojí řídit automobil, také více vyhýbají komplikovaným situacím na silnicích, což se dalo předpokládat. Nejistí řidiči se tak sice ubrání pro ně stresujícím situacím, ale svého strachu se tím nezbaví a své řidičské schopnosti tím nezdokonalí.

2.1.5 Shrnutí

Celkově lze říci, že kategorie svobodných a mladých řidičů neovlivňuje dopravní situace o moc negativněji, než ostatní řidiči. Nebylo prokázáno zvýšené riziko u řízení pod vlivem alkoholu, telefonování za volantem ani u rozčilování se na ostatní řidiče. Dokonce ani strach z řízení se u svobodných a mladých řidičů nevyskytuje častěji, než u ostatních.

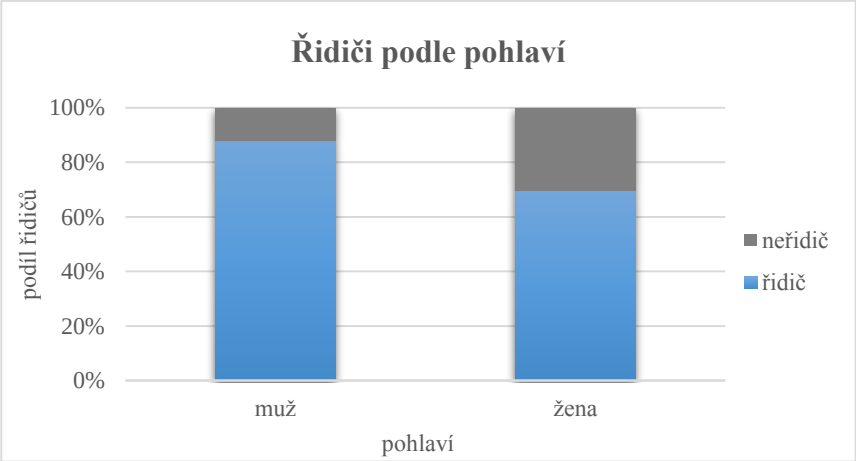
Zvýšené riziko spočívá zejména v telefonování a psaní SMS za volantem, které se rozmáhá více mezi mladými, a dále pak v řízení pod vlivem alkoholu, které je bližší svobodným mužům, než ženám, stejně jako rychlá jízda.

2.2 Genderová problematika u řidičů

Druhou oblastí, kterou se tato práce zabývá, je genderová problematika. Je ve všeobecném povědomí, že muži a ženy se za volantem chovají odlišně. Ačkoli se s postupem času rozdílnosti pomalu stírají, stále platí, že za volant sedají častěji muži, kteří ale také více riskují. Jak uvádí statistiky BESIP – Ministerstvo dopravy, muži vykazují přibližně dvojnásobnou míru nehodovosti na tisíc řidičů, přičemž svou první nehodu absolvují v mladším věku, než ženy. Mezi příčiny spadá velmi často rychlá jízda a menší respekt k pravidlům silničního provozu. Dalším příčinou je také řízení pod vlivem alkoholu nebo drog, s čímž mají problémy především muži, jak vyplývá ze závěrů Evropského projektu DRUID.[5]

Na základě těchto informací je v této kapitole zkoumáno odlišné chování mužů a žen za volantem. Pro základní seznámení s výběrem z pohledu pohlaví jsou zde následující grafy. Na obrázku 1.3 lze vidět, že je ve výběru více mužů řidičů, zatímco u neřidičů je větší podíl žen. Již z tohoto grafu by se dalo konstatovat, že za volant sedají více muži, zatímco značnou část neřidičů ve výběru tvoří ženy.

[5] BESIP – Ministerstvo dopravy (2014). *Muži a ženy – základní statistické ukazatele ve formě komentovaných grafů* [online]. s. 22.



Obrázek 1.3 Podíly řidičů a neřidičů podle pohlaví

2.2.1 Strach z řízení - neřidiči

V dotazníku se vyskytovaly také otázky pro neřidiče, v nichž měli odpovídat, z jakých důvodů automobil neřídí. Jak lze snadno dopočítat z tabulky 2.20, téměř 40 % žen, které neřídí, tak činí v důsledku toho, že nemají řidičský průkaz, a u mužských neřidičů je to dokonce téměř polovina. Z toho plyne, že pro neřízení mají ženy často jiné důvody.

Zajímavější výsledkem ale je, že 26 % žen, které se označily za neřidiče, sice řidičský průkaz má, ale bojí se řídit, zatímco u mužů se jedná pouze o 15 % takových. Konečně respondenti neřidiči, kteří řidičský průkaz mají a řídit se nebojí, uváděli často jako příčinu zdravotní důvody a několik žen pouze konstatovalo, že vždy řídí jejich manžel.

Tabulka 2.20 Trojrozměrná kontingenční tabulka – pohlaví, strach, bez řidičského průkazu

Z jakých důvodů neřídíte automobil?			Nemám řidičský průkaz		Celkem
			Ne	Ano	
Muž	Bojím se	Ne	15	18	33
		Ano	6	1	7
		Celkem	21	19	40
Žena	Bojím se	Ne	33	30	63
		Ano	25	8	33
		Celkem	58	38	96
Celkem			79	57	136

Pro určení závislosti mezi těmito třemi proměnnými jsou nejprve porovnána podmíněná a marginální rozdělení, která se v tomto případě téměř shodují, ale marginální poměr šancí je různý od jedné, což značí asociaci dvou proměnných bez vlivu veličiny třetí. Ke zjištění konkrétních vztahů jsou podle (1.42) vypočteny párové korelační koeficienty r_{xy} = -0,254, r_{yz} = -0,073 a r_{xz} = 0,169 za předpokladu, že *x* označuje strach z řízení, *y* značí řidičský průkaz a *z* pohlaví.

Dále jsou určeny parciální korelační koeficienty, tedy vždy bez vlivu třetí proměnné jako $r_{xy.z}$ = -0,246, $r_{yz.x}$ = -0,032 a $r_{xz.y}$ = 0,156. Z výsledků vyplývá pozitivní závislost mezi strachem z řízení a pohlavím respondenta a také negativní závislost mezi strachem z řízení a nevlastněním řidičského průkazu. Poměry šancí zvlášť pro muže a ženy vycházejí poměrně odlišně, proto není vypočítán společný MH poměr šancí.

Strach z řízení byl sice již zkoumán v kapitole o svobodných řidičích, ale ještě ne z hlediska pohlaví. V tabulce 2.21 jsou uvedeny absolutní četnosti vzniklé z tabulky 2.20 eliminováním vlivu proměnné *y*.

Tabulka 2.21 Čtyřpolní tabulka – pohlaví, strach z řízení

		Bojím se		Celkem
		Ne	Ano	
Pohlaví	Muž	33	7	40
	Žena	63	33	96
	Celkem	96	40	136

Po provedení chí-kvadrát testu je zjištěno, že testové kritérium X^2 = 3,873 je nepatrně vyšší, než hodnota kvantilu $\chi^2_{0,95}(1)$ = 3,841, takže lze na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti mezi strachem z řízení a pohlavím. Test může být taktéž proveden pomocí věrohodnostního poměru podle (1.8), kdy G^2 = 4,128 také konverguje k chí-kvadrát rozdělení se stejným počtem stupňů volnosti. I v tomto případě je testové kritérium vyšší, než příslušný kvantil.

Vypočtené míry intenzity závislosti se pohybují blízko párového korelačního koeficientu r_{xz} = 0,169, což ukazuje značnou závislost. K názornějšímu vyjádření této závislosti je vhodné vyčíslit šance a následně také poměr těchto šancí, přičemž vše je bráno ke skupině respondentů, kteří se označili za neřidiče. Šance, že neřídí muž, protože se za volantem bojí, je přibližně 1:5, zatímco šance, že neřídí žena z důvodu strachu, je více než 1:2. Pokud tedy někdo neřídí, protože se bojí, bude to s šancí 5:2 žena.

Je možno říci, že strach tvoří příčinu neřízení automobilu u žen až 2,5krát častěji, než u mužů. Většina mužů automobil řídí, a ti, kteří se mezi ně neřadí, tak mohou činit například z důvodu nevlastnění vozidla, ze zdravotních důvodů a pár mužských respondentů dokonce uvedlo, že má zákaz řízení. Již z pohledu neřidičů je pozorovatelné, že se zde vyskytuje odlišné chování u mužů a žen. Následující podkapitoly jsou zaměřeny především na skupinu řidičů.

2.2.2 Řízení jako obživa

Denně za volant sedá velká spousta lidí, a to zejména proto, aby se někam přepravili. Někteří to však mají jako povolání. Ať už se jedná o řidiče kamionů, autobusů nebo třeba řidiče taxi služby, všechny tato činnost živí. Dá se předpokládat, že se jedná zejména o mužskou profesi, a tedy že žen, které se řízením živí, bude minimum. Takové zkoumání je předmětem této podkapitoly. V tabulce 2.22 lze vidět, kolik respondentů, mužů a žen, kteří se označili za řidiče, se živí řízením motorového vozidla.

Tabulka 2.22 Čtyřpolní tabulka – pohlaví, řízení jako obživa

Co pro vás znamená automobil?		Pohlaví		Celkem
		Muž	Žena	
Moje obživa	Ne	267	212	479
	Ano	26	9	35
	Celkem	293	221	514

Podle chí-kvadrát testu o nezávislosti a testu věrohodnostním poměrem bylo zjištěno, že na 5% hladině významnosti lze zamítnout nulovou hypotézu o nezávislosti těchto dvou proměnných, neboť jsou testová kritéria $X^2 = 4{,}577$ a $G^2 = 4{,}827$ větší, než kvantil $\chi^2_{0,95}(1) = 3{,}841$. Nicméně vypočtené míry intenzity závislosti těchto dvou proměnných, které lze vidět v tabulce 2.23, ukazují na nepatrnou závislost.

Tabulka 2.23 Výpočty – koeficienty pro čtyřpolní tabulku – pohlaví, řízení jako obživa

C_P	C_P^*	$V = \phi$	τ_{xy}	U_{xy}	λ_{xy}
0,094	0,133	0,094	0,009	0,019	0,000

Symetrické míry intenzity závislosti vypočtené podle (1.20 – 1.26) nabývají vcelku nízké hodnoty, což lze nejlépe pozorovat z upraveného Pearsonova koeficientu kontingence C_P^*, který nabývá hodnot od 0 do 1. Hodnota 0,133 tedy vyjadřuje poměrně nízkou závislost mezi pohlavím a povoláním zahrnující řízení. Velikost jednostranné závislosti řízení jako obživy na pohlaví vyjadřují nesymetrické koeficienty, které jsou taktéž uvedeny v tabulce 2.23. Tyto koeficienty jsou poměrně blízko nuly, a dalo by se tedy říci, že tento způsob obživy až tak výrazně na pohlaví nezávisí. Lambda nabývá nuly, jelikož jsou modální kategorie v obou případech zastoupeny ve stejném sloupci.

Dále jsou vypočteny šance, kterými lze vyjádřit závislost mezi proměnnými poněkud názornějším způsobem. Šance, že se bude řízením živit muž, je přibližně téměř 1:10, zatímco šance, že se bude řízením živit žena, je až 1:25, což je značný rozdíl. Z poměru těchto šancí pak plyne šance 23:10, že člověk, který se řízením motorového vozidla živí, bude muž. Ačkoli bylo ukázáno, že závislost mezi řízením jako obživou na pohlaví je poměrně nízká, přesto by se toto

povolání dalo označit za spíše mužské, neboť je možné se s 2,3krát vyšší šancí setkat s řidičem z povolání mužem, než se ženou.

2.2.3 Řízení pod vlivem alkoholu

Již v kapitole týkající se svobodných respondentů byl porovnáván rozdíl v řízení pod vlivem alkoholu u svobodných mužů a žen. V této podkapitole jsou bráni v úvahu všichni respondenti. Nejprve byla do tabulky 2.24 zakomponována navíc veličina, která určuje, zda respondent má nebo nemá děti. Dalo by se předpokládat, že řidiči s dětmi budou méně riskovat.

Tabulka 2.24 Trojrozměrná kontingenční tabulka – děti, alkohol, pohlaví

Máte děti?			Pohlaví		Celkem
			Muž	Žena	
Ne	Alkohol	Ano	67	19	86
		Ne	148	155	303
		Celkem	215	174	389
Ano	Alkohol	Ano	39	18	57
		Ne	75	122	197
		Celkem	114	140	254
Celkem			329	314	643

Podmíněná rozdělení se přibližně shodují s marginálním, ale poměry šancí jsou různé od jedné, z čehož vyplývá, že se zde vyskytuje závislost alkoholu na pohlaví, ale obě tyto veličiny jsou zároveň nezávislé na proměnné týkající se dětí, což lze vyčíst také z parciálních korelačních koeficientů. Nejprve jsou vypočteny párové korelační koeficienty podle (1.41), které vycházejí r_{xy} = 0,246, r_{yz} = 0,102 a r_{xz} = -0,004, za předpokladu, že *x* označuje alkohol za volantem, *y* značí pohlaví a *z* značí děti. Následně lze podle (1.42) dopočítat parciální korelační koeficienty $r_{xy.z}$ = 0,247, $r_{yz.x}$ = 0,106 a $r_{xz.y}$ = -0,030, které vyjadřují závislost dvou proměnných bez vlivu třetí, uvedené v dolním indexu za tečkou.

Tyto výsledky potvrzují závislost mezi alkoholem a pohlavím, která je ale nezávislá na počtu dětí. To může být způsobeno tím, že se otázka řízení pod vlivem alkoholu neptá pouze na dobu, kdy má respondent děti, ale na celý život. Takže respondent mohl například řídit pod vlivem alkoholu v mládí, ale po narození dětí už takto neriskoval, což se ale ve výsledných statistikách neprojeví.

Po eliminování třetí veličiny je získána čtyřpolní tabulka 2.25, z níž lze vyčíst, že z respondentů, kteří řídili pod vlivem alkoholu, je 74 % mužů.

Tabulka 2.25 Kontingenční tabulka – alkohol, pohlaví

		Pohlaví		Celkem
		Muž	Žena	
Řídil/a jste někdy pod vlivem alkoholu?	Ano	106	37	143
	Ne	223	277	500
	Celkem	329	314	643

Po provedení chí-kvadrát testu bylo zjištěno, že testové kritérium X^2 = 38,80 je mnohem vyšší, než hodnota kvantilu $\chi^2_{0,95}(1)$ = 3,841, takže lze na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti proměnných. Dále je předmětem zájmu zjistit, zda opravdu řídí pod vlivem alkoholu spíše muži, jak bylo předpokládáno, a v jakém poměru. K tomu lze využít šance.

Šance, že muž někdy řídil pod vlivem alkoholu, je přibližně 1:2, zatímco u žen je šance řízení pod vlivem alkoholu menší než 1:7. Po vypočtení poměru šancí lze říci, že šance, že pod vlivem alkoholu řídí muž, oproti šanci, že pod vlivem řídí žena, je přibližně 7:2. Tento výsledek potvrzuje prvotní předpoklad, že muži s alkoholem riskují častěji, než ženy, a to dokonce 3,5 krát.

2.2.4 Telefonování a psaní SMS za volantem

V kapitole týkající se svobodných řidičů bylo již telefonování a psaní SMS za volantem zkoumáno, ale ne z hlediska pohlaví. V tabulce 2.26 jsou uvedeny absolutní četnosti odpovědí respondentů na otázky ohledně telefonování a psaní zpráv za volantem, rozčleněné podle pohlaví.

Tabulka 2.26 Trojrozměrná kontingenční tabulka – pohlaví, telefonování a psaní SMS

			Psaní SMS za volantem		Celkem
			Ne	Ano	
Muž	Telefonování za volantem	Ne	123	16	139
		Ano	74	80	154
		Celkem	197	96	293
Žena	Telefonování za volantem	Ne	103	8	111
		Ano	59	51	110
		Celkem	162	59	221
Celkem			359	155	514

Podmíněná rozdělení se téměř shodují s marginálním, ale podmíněný poměr šancí není roven jedné, což ukazuje na závislost dvou proměnných bez vlivu proměnné třetí. Vypočtením parciálních korelačních koeficientů je možné zjistit, zda se jedná pouze o závislost mezi telefonováním a psaním SMS za volantem, která byla již zkoumána v kapitole o svobodných řidičích, nebo zda se zde vyskytuje i jiná závislost. Jestliže *x* označuje telefonování za volantem,

y značí psaní SMS a z značí pohlaví, pak jsou parciální korelační koeficienty vypočtené podle (1.42) $r_{xy.z} = 0{,}435$, $r_{yz.x} = -0{,}059$ a $r_{xz.y} = -0{,}001$. Z těchto hodnot lze vyčíst středně silná závislost mezi telefonováním a psaním SMS za volantem, která již byla zkoumána výše, a dále možná velmi malá negativní závislost mezi pohlavím a telefonováním, případně nezávislost mezi pohlavím a psaním SMS za volantem.

U telefonování za volantem a pohlavím byl ještě proveden chí-kvadrát test o nezávislosti a test věrohodnostním poměrem, v případě, že by se zde velmi malá závislost přeci jen vyskytovala. Testová kritéria ale nabývala shodně hodnoty $X^2 = G^2 = 0{,}391$, která není větší, než kvantil $\chi^2_{0,95}(1) = 3{,}84$, a proto není možné na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti mezi telefonováním za volantem a pohlavím.

Závislost mezi pohlavím a telefonováním za volantem nebyla prokázána, a stejně tak nebyla prokázána ani závislost mezi pohlavím a psaním SMS při řízení. Nelze tedy říci, že by muži nebo ženy riskovali telefonováním a psaním zpráv více, než ti druzí.

2.2.5 Zpívání za volantem

Ačkoli se jedná spíše o zajímavost, která neohrožuje situace na silnicích tak, jako alkohol nebo rychlá jízda, bylo do analýzy zařazeno i zpívání při řízení. Dalo by se předpokládat, že ženy budou za volantem zpívat častěji, než muži a cílem je potvrdit, že to ze získaných dat také vyplývá. V tabulce 2.27 jsou zobrazeny absolutní četnosti odpovědí na otázku zpívání při řízení, rozčleněné podle pohlaví.

Tabulka 2.27 Kontingenční tabulka – zpívání, pohlaví

		Pohlaví		
		Muž	Žena	Celkem
Jak často zpíváte při řízení?	Nikdy	78	20	98
	Výjimečně	99	38	137
	Občas	80	85	165
	Často	30	62	92
	Téměř pokaždé	6	16	22
	Celkem	293	221	514

Nejprve je proveden chí-kvadrát test o nezávislosti proměnných. Testové kritérium, které je vypočtené podle (1.6) $X^2 = 68{,}57$ nabývá výrazně vyšší hodnoty, než $\chi^2_{0,95}(4) = 9{,}488$, a proto lze na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti těchto proměnných. Stejného závěru by bylo dosaženo použitím testu věrohodnostním poměrem, jehož testové kritérium je $G^2 = 70{,}93$.

Tabulka 2.28 Výpočty – koeficienty nominálních proměnných – zpívání, pohlaví

C_P	C_P^*	V	ϕ	τ_{xy}	U_{xy}	λ_{xy}
0,343	0,485	0,365	0,365	0,034	0,047	0,054

V tabulce 2.28 jsou uvedeny nejprve symetrické míry intenzity závislosti, které vyjadřují oboustrannou závislost mezi zpíváním za volantem a pohlavím. Upravený Pearsonův koeficient kontingence, který nabývá hodnoty od 0 do 1, se zde rovná téměř 0,5, a tedy lze označit oboustrannou závislost za středně silnou. Jednostranná závislost je vyjádřena pomocí nesymetrických koeficientů, z nichž lze odvodit poměrně malou závislost zpívání za volantem na pohlaví. Zkoumání opačné závislosti by nemělo věcnou interpretaci.

To však stále nevypovídá o tom, zda si za volantem častěji zpívají muži či ženy. Vypočtením šancí lze dospět k názornější interpretaci. Jsou-li bráni v úvahu ti respondenti, kteří odpověděli, že si nikdy nezpívají za volantem, je u nich šance přibližně 1:4, že mezi nimi bude žena. Na druhou stranu pokud jsou zkoumáni ti, kteří si za volantem zpívají často, je u nich šance více než 2:1, že je mezi nimi žena. Tím je prokázáno, že si ženy za volantem zpívají častěji, než muži.

Pro zajímavost bylo ještě zjišťováno, zda ženy zpívají často s puštěným rádiem, či bez něj. V následující tabulce 2.29 jsou uvedeny absolutní četnosti odpovědí respondentek.

Tabulka 2.29 Kontingenční tabulka – Uveďte, jak často se věnujete následujícím činnostem?

Zpívání	Poslech rádia					Celkem
	Nikdy	Výjimečně	Občas	Často	Téměř pokaždé	
Nikdy	2	0	2	4	12	20
Výjimečně	2	2	6	12	16	38
Občas	0	4	12	25	44	85
Často	0	1	5	13	43	62
Téměř pokaždé	0	0	0	1	15	16
Celkem	4	7	25	55	130	221

Nejprve byl podle (1.9) vypočten Spearmanův koeficient korelace pořadí $r_S = 0{,}239$, který značí slabou pozitivní lineární závislost mezi těmito proměnnými, která je prokázána otestováním nulovosti Spearmanova koeficientu. Podle (1.14) je vypočteno testové kritérium $t = 3{,}649$, které nabývá vyšší hodnoty než kvantil $u_{0,975} = 1{,}96$, tudíž lze na 5% hladině významnosti zamítnout nulovou hypotézu o nezávislosti proměnných.

Následně byly vypočteny míry intenzity závislosti pro ordinální veličiny podle (1.32 – 1.35), jejichž výsledky lze vidět v tabulce 2.30. Všechny koeficienty jsou kladné, což odpovídá pozitivní lineární závislosti vzhledem k uspořádání v tabulce. S rostoucími četnostmi kategorií jedné proměnné tedy rostou četnosti kategorie druhé proměnné.

Tabulka 2.30 Výpočty – koeficienty ordinálních proměnných – zpívání, poslech rádia

Γ	τ_B	τ_C	d_{yx}	d_{xy}
0,323	0,210	0,171	0,187	0,236

První tři koeficienty uvedené v tabulce 2.24 určují oboustrannou závislost těchto proměnných, přičemž by mezi ně patřilo i symetrické Somersovo *d*, které však nabývá stejné hodnoty, jako Kendallovo τ_B. Z vypočtených koeficientů vyplývá, že závislost mezi zpíváním za volantem a posloucháním rádia je poměrně slabá.

Zajímavé je v tomto případě porovnání dvou nesymetrických Somersových koeficientů *d*, které určují jednostrannou závislost veličin. Z hodnot těchto koeficientů vyplývá, že je více závislé zpívání na poslechu rádia, než poslech rádia na zpívání, což dává smysl. Pokud bude žena při řízení poslouchat rádio, pak si spíše začne i zpívat, než naopak.

2.2.6 Shrnutí

Odlišnosti v chování mužů a žen se vyskytují ve všech činnostech, zejména pak při řízení vozidel. Muži řídí častěji, než ženy, které se podle výsledků zkoumání častěji bojí za volantem. Dále výsledky potvrdily, že řízení motorových vozidel je spíše mužské povolání, přestože rozdíly mezi řidiči a řidičkami z profese nebyly tak markantní.

Řízení pod vlivem alkoholu bylo tentokrát zkoumáno nejen pro svobodné řidiče, ale pro všechny muže a ženy, přičemž vyšlo najevo, že muži řídí po požití alkoholu až 3,5krát častěji, než ženy. Naproti tomu u telefonování a psaní SMS za volantem nebyla prokázána závislost na pohlaví.

Pro zajímavost bylo ještě zjištěno, že ženy si za volantem zpívají mnohem častěji, než muži, a byla prokázána lineární závislost mezi zpíváním a posloucháním rádia při řízení. Celkově tedy riziko spočívá zejména v řízení pod vlivem alkoholu, které je častější u mužů, dále pak v telefonování a psaní zpráv za volantem, které hrozí u obou pohlaví.

Závěr

Tato bakalářská práce se zabývala analýzou dat z dotazníkového šetření řidičů pomocí kontingenčních tabulek. Náplní první části práce bylo popsání základních metod analýzy kategoriálních dat. Byly zde zmíněny testy o nezávislosti proměnných, šance a poměr šancí, a také způsob měření intenzity závislosti pomocí různých koeficientů pro nominální a ordinální proměnné. Metody analýzy byly uvedeny zejména pro dvourozměrné kontingenční tabulky, ale byl zde vysvětlen také úvod do problematiky vícerozměrného třídění.

Cílem práce byla analýza chování svobodných řidičů a odlišností mezi muži a ženami v řízení, a to zejména v případě činností, které mohou zvyšovat riziko vzniku nehod na silnicích. Získané výsledky byly otestovány a zobecněny na celou populaci.

Hlavní částí práce byla část praktická, kde byly metody analýzy kontingenčních tabulek aplikovány na reálná data týkající se řidičů v České republice. Nejprve byla zkoumána kategorie svobodných a mladých řidičů, a to zejména v důsledku předpokládaných menších zkušeností a menší zodpovědnosti oproti ostatním. Z výsledků analýzy vyšlo najevo, že řízení pod vlivem alkoholu nezávisí na rodinném stavu řidičů, a tedy nelze ani říci, že by svobodní mladí řidiči častěji podstupovali riziko řízením v podnapilém stavu. U svobodných řidičů ale bylo zjištěno, že čím více věří svým řidičským dovednostem, tím více telefonují za volantem, přičemž bylo potvrzeno, že mladí a svobodní za volantem také častěji píší SMS, než ostatní.

Dále byla prokázána malá závislost mezi rozčilováním se na ostatní řidiče a rodinným stavem, přičemž svobodní řidiči, kteří se častěji rozčilují při řízení, také častěji upozorňují ostatní řidiče například na nerozsvícená světla. Jako poslední byl u svobodných řidičů zkoumán strach z řízení, u něhož ale nebyla prokázána závislost na rodinném stavu. Bylo však zjištěno, že čím více se svobodní řidiči řízení bojí, tím častěji se vyhýbají větším městům nebo komplikovaným křižovatkám, čímž svůj strach nezmenšují.

V druhé části byly zkoumány odlišnosti mezi muži a ženami ve vztahu k řízení. Z výsledků vyplynulo, že se ženy za volantem bojí častěji, než muži, kteří sedají za volant častěji. Výsledky dále potvrdily předpoklad, že je řízení motorových vozidel převážně mužské povolání, přestože rozdíly mezi řidiči a řidičkami z profese nebyly tak markantní.

Řízení pod vlivem alkoholu bylo zkoumáno nejen pro svobodné řidiče, ale pro všechny muže a ženy, přičemž vyšlo najevo, že muži řídí po požití alkoholu až 3,5krát častěji, než ženy. Naproti tomu u telefonování a psaní SMS za volantem nebyla prokázána závislost na pohlaví.

Pro zajímavost bylo ještě zjištěno, že ženy si za volantem zpívají mnohem častěji, než muži, a byla prokázána lineární závislost mezi zpíváním a posloucháním rádia při řízení.

Tyto výsledky přináší subjektivnější pohled na chování řidičů v porovnání se statistikami nehodovosti, neboť vychází z dotazníků, které vyplňovali sami řidiči podle svého uvážení. Je zde porovnání kategorie svobodných řidičů s ostatními, což obvykle zkoumáno nebývá, a dále pak odlišnosti mezi muži a ženami, které vnáší nové poznatky do genderové problematiky.

Na základě těchto výsledků lze považovat cíl bakalářské práce za splněný. Vzhledem k velkému rozsahu datového souboru nebylo možné analyzovat všechny veličiny, tudíž by se v budoucnu práce dala rozšířit ještě o analýzu dalších proměnných nebo se zaměřit na jiné charakteristické rysy řidičů. Například bych navrhovala zkoumat chování řidičů podle vzdělání, výše příjmu či typu vozidla. Přínosné by také bylo původní dotazník rozšířit o dotazy týkající se nehodovosti.

Přehled literatury

AGRESTI, Alan. *An Introduction to Categorical Data Analysis*. 2. vyd. New Jersey: John Wiley & Sons, Inc., 2007. ISBN: 978-0-471-22618-5.

HINDLS, Richard a kol. *Statistika pro ekonomy*. 7. vyd. Praha: Professional Publishing, 2006. ISBN: 80-86946-16-9

PECÁKOVÁ, Iva. *Statistika v terénních průzkumech*. 2. vyd. Praha: Professional Publishing, 2011. ISBN: 978-80-7431-039-3

ŘEZANKOVÁ, Hana. *Analýza dat z dotazníkových šetření*. 1. vyd. Praha: Professional Publishing, 2007. ISBN: 978-80-86946-49-8

Internetové zdroje

BESIP – Ministerstvo dopravy ČR. *Muži a ženy – základní statistické ukazatele ve formě komentovaných grafů*, 2014. [online]

Data Collect, s.r.o. – omnibus, http://www.omnibusy.cz/omnibus.php

KAMENICKÝ, J. *Dopravní nehodovost a její důsledky v ČR v dlouhodobém pohledu* [online], ČSÚ, 2014.

Seznam tabulek

Tabulka 1.1 Kontingenční tabulka – symbolika ... 3
Tabulka 1.2 Vícerozměrná kontingenční tabulka ... 14
Tabulka 2.1 Kontingenční tabulka – rodinný stav, alkohol ... 19
Tabulka 2.2 Kontingenční tabulka – pohlaví, alkohol ... 20
Tabulka 2.3 Výpočty – symetrické koeficienty – pohlaví, alkohol ... 20
Tabulka 2.4 Výpočty - nesymetrické koeficienty – pohlaví, alkohol ... 21
Tabulka 2.5 Trojrozměrná kontingenční tabulka – alkohol, rychlost, rodinný stav ... 21
Tabulka 2.6 Čtyřpolní tabulka – alkohol, rychlost ... 22
Tabulka 2.7 Výpočty – koeficienty pro čtyřpolní tabulku – alkohol, rychlost ... 23
Tabulka 2.8 Kontingenční tabulka – Jak byste charakterizoval/a své řidičské schopnosti, telefonování ... 23
Tabulka 2.9 Počty dvojic ordinálních proměnných – řidičské schopnosti, telefonování ... 24
Tabulka 2.10 Výpočty – koeficienty ordinálních proměnných – řidičské schopnosti, telefonování ... 24
Tabulka 2.11 Trojrozměrná kontingenční tabulka – rodinný stav, telefonování, SMS ... 25
Tabulka 2.12 Výpočty – koeficienty nominálních proměnných – telefonování, psaní SMS ... 26
Tabulka 2.13 Kontingenční tabulka – rozčilování za volantem, rodinný stav ... 27
Tabulka 2.14 Výpočty – koeficienty nominálních proměnných rozčilování se a rodinný stav ... 27
Tabulka 2.15 Kontingenční tabulka – rozčilování se a upozorňování ostatních řidičů ... 28
Tabulka 2.16 Výpočty – koeficienty – rozčilování se a upozorňování na nerozsvícená světla ... 28
Tabulka 2.17 Kontingenční tabulka – strach za volantem, rodinný stav ... 29
Tabulka 2.18 Kontingenční tabulka – strach z řízení, vyhýbání se městu ... 29
Tabulka 2.19 Výpočty – koeficienty ordinálních proměnných – strach z řízení, vyhýbání se městu ... 30
Tabulka 2.20 Trojrozměrná kontingenční tabulka – pohlaví, strach, bez řidičského průkazu ... 31
Tabulka 2.21 Čtyřpolní tabulka – pohlaví, strach z řízení ... 32
Tabulka 2.22 Čtyřpolní tabulka – pohlaví, řízení jako obživa ... 33
Tabulka 2.23 Výpočty – koeficienty pro čtyřpolní tabulku – pohlaví, řízení jako obživa ... 33
Tabulka 2.24 Trojrozměrná kontingenční tabulka – děti, alkohol, pohlaví ... 34
Tabulka 2.25 Kontingenční tabulka – alkohol, pohlaví ... 35
Tabulka 2.26 Trojrozměrná kontingenční tabulka – pohlaví, telefonování a psaní SMS ... 35
Tabulka 2.27 Kontingenční tabulka – zpívání, pohlaví ... 36
Tabulka 2.28 Výpočty – koeficienty nominálních proměnných – zpívání, pohlaví ... 37
Tabulka 2.29 Kontingenční tabulka – Uveďte, jak často se věnujete následujícím činnostem? ... 37
Tabulka 2.30 Výpočty – koeficienty ordinálních proměnných – zpívání, poslech rádia ... 38

Seznam obrázků

Obrázek 1.1 Výsečový graf – rodinný stav ... 17
Obrázek 1.2 Histogram – věk svobodných respondentů ... 18
Obrázek 1.3 Podíly řidičů a neřidičů podle pohlaví ... 31