Z této věty vyplývá, že pro velká n lze rozdělení pravděpodobnosti náhodné veličiny $\sum_{j=1}^{n} X_j$, $j = 1, 2, \ldots$, aproximovat rozdělením $N(n\mu, n\sigma^2)$ a rozdělení náhodné veličiny $\sum_{j=1}^{n} X_j / n$ lze aproximovat rozdělením $N(\mu, \sigma^2/n)$. Důležité je, že tak lze učinit za poměrně obecných předpokladů a bez znalosti původního rozdělení veličin X_j.

Moivreova-Laplaceova věta

Speciálním případem Lindebergovy-Lévyho věty je následující věta: Nechť $\{X_n\}$ je posloupnost vzájemně nezávislých náhodných veličin s alternativním rozdělením se stejným parametrem π, $0 < \pi < 1$. Pak posloupnost $\{Y_n\}$, kde

$$Y_n = \frac{1}{[n\pi(1-\pi)]^{1/2}} \left(\sum_{j=1}^{n} X_j - n\pi \right), \tag{2.167}$$

konverguje v distribuci k rozdělení $N(0, 1)$.

Příklad 2.15

Na základě minulých pozorování víme, že průměrná hmotnost balíku zasílaného poštou je 5 kg. Hmotnost zasílaného balíku můžeme považovat za náhodnou veličinu se střední hodnotou 5 kg a se směrodatnou odchylkou 1,5 kg (ta byla rovněž spočítána na základě minulých pozorování). Jaká je pravděpodobnost, že 100 zasílaných balíků, které se nakládají do poštovního auta, převýší svou hmotností užitečné zatížení nákladního prostoru tohoto automobilu, které činí 550 kg?

Řešení

Hmotnost i-tého balíku zasílaného poštou můžeme považovat za náhodnou veličinu X_i pro $i = 1, 2, \ldots, 100$ se střední hodnotou $\mu = 5$ kg a směrodatnou odchylkou $\sigma = 1{,}5$ kg. Součet hmotnosti všech 100 balíků označme Y, tedy $Y = \sum_{i=1}^{100} X_i$. Hmotnosti jednotlivých balíků jsou nezávislé. Dále platí, že všechny veličiny X_i mají stejnou střední hodnotu a rozptyl. Jsou tedy splněny podmínky použití centrální limitní věty a lze psát

$$Y = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2).$$

Po dosazení obdržíme

$$Y = \sum_{i=1}^{100} X_i \sim N(100 \cdot 5, 100 \cdot 2{,}25), \text{ tedy } n = 100.$$

Hledaná pravděpodobnost je pak rovna

$$P(Y > 550) = 1 - P(Y \le 550) = 1 - P\left(\frac{Y - 500}{15} < \frac{550 - 500}{15} \right) =$$

$$= 1 - P(U \le 3{,}333) = 1 - F(3{,}333) = 1 - 0{,}99957 = 0{,}0004.$$

■

3 ANALÝZA DAT Z VÝBĚROVÝCH ŠETŘENÍ

V této části použijeme znalosti z předchozích dvou kapitol věnovaných popisné statistice a teorii pravděpodobnosti k tomu, abychom se pokusili usuzovat na základě omezené informace obsažené v nějakém vzorku na celou populaci, která je cílem zkoumání. Takovými postupy se zabývá matematická statistika, která používá metody teorie pravděpodobnosti při analýze dat. Metody a postupy, které umožňují zobecňovat výsledky z části na celek, se nazývají statistickou indukcí.

Nejprve se budeme věnovat problémům spojeným s pořizováním dat, ze kterých by bylo možné činit spolehlivé závěry, a dále představíme základní metody zobecňování, kterými jsou bodový a intervalový odhad a testování statistických hypotéz.

3.1 Výběrová šetření

Žádná statistická metoda nemůže poskytnout hodnotné výsledky, jsou-li špatná data, se kterými pracujeme. O kvalitě dat rozhoduje již fáze statistického šetření, v rámci kterého je získáváme. První podstatnou okolností při jeho přípravě je rozsah **statistického souboru**, který má být zkoumán (používáme také název **základní soubor**, **základní populace** nebo jen **populace**). Může to být soubor všech domácností v České republice, soubor všech voličů, kteří se mohou zúčastnit určitých voleb, nebo soubor všech příjemců nějaké dávky.

Ve **vyčerpávajícím zjišťování** se prošetřují veškeré jednotky základního souboru. Uvažujme rozsáhlý soubor – řekněme, že ho tvoří například všichni obyvatelé nějakého území. Realizace takového statistického šetření bude vyžadovat náročnou organizaci, množství pracovníků, značné finanční i časové náklady. I zpracování jeho výsledků může zabrat relativně dlouhé období. Na druhé straně dobře provedené vyčerpávající zjišťování může poskytnout detailní informace o každé jednotce v základním souboru a přesné charakteristiky sledovaných veličin. Tyto jeho přednosti však zmíněnou náročnost vyčerpávajícího šetření málokdy vyváží, a tak nebývá realizováno často. Jako příklad uveďme sčítání lidu, domů a bytů, které náš stát a jeho statistické orgány realizují ve zhruba desetiletých intervalech.

Provedení vyčerpávajícího zjišťování však mohou bránit i jiné okolnosti. V průmyslové výrobě mohou určité zkoušky výrobků při ověřování jejich vlastností – například jejich odolnosti při určité zátěži – vést až k jejich znehodnocení, a tím k vyřazení z dalšího použití. Při dotazování zákazníků obchodu jejich kompletní soubor jednoduše ani nemáme k dispozici.

V takovém případě lze přistoupit k nevyčerpávajícímu, **výběrovému zjišťování**. To znamená, že prošetřeny jsou pouze některé jednotky, které byly ze základního souboru vybrány. V ekonomické praxi je sledován cenový vývoj nebo vývoj životních nákladů. Rozsáhlá komoditní skladba trhu a liberální tvorba cen vytvářejí situace, kdy

stejný výrobek je prodáván za různé ceny často ve stovkách prodejen na území celého státu. Podobně je to s tisíci dalších produktů. Pak nezbývá než zjišťovat tzv. reprezentanty ve vybraných prodejních místech a z vývoje jejich cen usuzovat na celkovou dynamiku cen. Stejně tak záznamy o výdajích domácností (kdy v rámci České republiky hovoříme o milionech jednotek) lze vést pouze ve výběrovém souboru. V sociologické oblasti se běžně setkáváme s výzkumy veřejného mínění či volebních preferencí, firmy zajímá poptávka jejich zákazníků, televizní stanice zkoumají sledovanost svých pořadů atd. Společným jmenovatelem všech těchto činností je výběrové setření.

Informace získané z výběrového zjišťování se poté obvykle zobecňují na celý základní soubor. Takový postup je ostatně celkem obvyklý i v běžném životě, kdy činíme závěry na základě jen kusých informací, odhadujeme určité skutečnosti nebo ověřujeme tvrzení o nich vyslovená. Úroveň takových úsudků je pochopitelně různá. Od statistických výběrových šetření však očekáváme, že jejich zobecnění na základní soubor budou nejen možná, ale že bude možné získat představu také o jejich kvalitě, kontrolovat ji a ovlivnit. To ovšem vyžaduje, abychom při realizaci výběrového šetření nepostupovali libovolně, ale abychom respektovali určitá pravidla, která zobecnění výběrových informací na základní soubor umožní.

Shrneme-li to, co jsme zatím o výběrovém šetření řekli, lze konstatovat, že ze základního souboru o velkém rozsahu (označíme ho N) vybereme jeho část, **výběrový soubor** (nebo také **vzorek**) o rozsahu n. Ve vzorku provedeme statistické šetření. Jeho výsledky zpracujeme a následně využijeme k provedení úsudků o základním souboru.

Způsob pořízení výběrového souboru je pro možnosti tohoto usuzování klíčový. Získat představu o kvalitě úsudků o zkoumané populaci, provedených na základě výběrových informací, umožňují pravděpodobnostní nebo také náhodné výběry jednotek. Data, která nebyla pořízena náhodným výběrem, takovou možnost nedávají.

Pravděpodobnostní (také **náhodný**) **výběr** znamená, že každá jednotka populace má známou pravděpodobnost zahrnutí do výběrového souboru. Přímý výběr jednotek se stejnými pravděpodobnostmi vybrání, který zajišťuje všem jednotkám také stejné pravděpodobnosti zahrnutí do vzorku a rovněž všem vzorkům o stanoveném rozsahu n stejnou pravděpodobnost jejich vzniku, je obvykle označován jako **prostý náhodný výběr**. Pokud má vybrání jednotky v kterémkoliv tahu probíhat za stále stejných podmínek, musí být jednotlivé tahy nezávislé na předchozích pokusech, což zajistí vrácení každé vybrané jednotky zpět do základního souboru (výběr s vracením nebo také s opakováním).

Uvažujme rozsáhlý, konečný základní soubor N domácností. Sledovanou veličinou budiž počet členů domácnosti, který označíme X. Náhodně (například na základě losu nebo nějakého generátoru náhodných čísel v softwaru) vybereme jednu domácnost a zjistíme, že počet jejích členů (náhodná veličina X_1) je například 3 (jedná se o hodnotu náhodné veličiny, tedy můžeme psát $x_1 = 3$). Je zřejmé, že pravděpodobnost, že první náhodně vybraná domácnost je tříčlenná, je dána tím, nakolik jsou tříčlenné domácnosti zastoupeny v základním souboru. Totéž platí i pro každou další možnou hodnotu této náhodné veličiny.

Kompletní rozdělení pravděpodobnosti náhodné veličiny *počet členů náhodně vybrané domácnosti* je tedy dáno rozdělením počtu členů domácnosti v základním souboru. V situaci velmi rozsáhlého, případně nekonečného základního souboru můžeme v této úvaze (podle zákona velkých čísel, který je vysvětlen v druhé kapitole) namísto rozdělení četností použít rozdělení pravděpodobnosti náhodné veličiny *počet členů domácnosti*. Říkáme pak, že pořizujeme **náhodný výběr z pravděpodobnostního rozdělení**.

Vrácením vybrané domácnosti zpět obnovíme výchozí stav v základním souboru a náhodně vybereme druhou domácnost. Počet jejích členů (náhodná veličina X_2) nabývá hodnoty x_2. Také rozdělení této náhodné veličiny je dáno rozdělením četností v základním souboru, a je tedy stejné jako rozdělení pravděpodobnosti náhodné veličiny X_1. Navíc výsledek druhého tahu nezávisí na výsledku tahu prvního, X_1 a X_2 jsou tedy nezávislé náhodné veličiny. Obdobné závěry platí i pro další náhodné veličiny $X_3, X_4, \ldots, X_n$, tedy počet členů třetí, čtvrté, až n-té náhodně vybrané domácnosti. Všechny veličiny mají shodné pravděpodobnostní rozdělení odpovídající rozdělení studovaného znaku v základním souboru a jsou nezávislé.

Popsaný postup, prostý náhodný výběr s vracením, je nejjednodušším typem pravděpodobnostního výběru. Zároveň je nejvýhodnější pro teoretické úvahy, na jejichž základě realizujeme úsudky o charakteristikách základního souboru i o jejich kvalitě.

Jednotkám v základním souboru mohou být přiřazeny i různé pravděpodobnosti vybrání, mají pak také různou šanci dostat se do výběru. Takový postup je volen, pokud pro hodnocení určitého problému nemají jednotky (řekněme firmy) stejný význam, neboť jsou například nestejně velké. Vyjádření míry odlišnosti jednotek, a tedy různých pravděpodobností jejich zahrnutí do výběru, může ovšem být v praxi dost komplikované.

Zajistit nezávislost náhodných veličin vracením jednotek do základního souboru v procesu jejich vybírání je ovšem velmi nepraktické a výběry jsou realizovány obvykle bez vracení. Tato skutečnost neznamená komplikaci pouze tehdy, představuje-li vzorek jen nepatrnou část rozsáhlého základního souboru, řekněme do 5 %, nebo pokud dokonce uvažujeme základní soubor nekonečně velký. V takovém případě jsou změny pravděpodobností vybrání jednotky v jednotlivých krocích výběru bez vracení natolik nepodstatné, že lze uvažovat teoreticky jednodušší výběr s vracením.

Pokud vzorek představuje významnější část základního souboru, je rozdíl mezi oběma výběrovými postupy zásadnější. Zejména v případě výběru s nestejnými pravděpodobnostmi vracení jednotek znamená, že do vzorku jsou častěji zahrnovány jednotky s největšími pravděpodobnostmi vybrání. Naštěstí je náhodný výběr bez vracení z hlediska kvality následných úsudků o základním souboru výhodnější než výběr s vracením. Při daném rozsahu totiž vzorek může zahrnovat větší počet různých výběrových jednotek než v případě opakovaného zařazení některých jednotek do výběru.

Realizace pravděpodobnostního výběru obvykle nepředstavuje fyzické vybírání jednotek. Probíhá na základě jejich seznamu, který se nazývá **opora výběru**. Jednotky jsou zde zastoupeny nějakým symbolem, nejčastěji pořadovým číslem, a vybírána jsou tato čísla.

Ale i náhodné vybírání – losování – například cestou očíslovaných lístků by v rozsáhlém souboru bylo technicky těžko proveditelné. V minulosti byly pro tento účel vytvářeny tabulky náhodných čísel, které obsahovaly číslice 0, 1, …, 9 uspořádané do sloupců tak, aby bylo možné je libovolně spojovat a vytvářet tak čísla o potřebném počtu míst. V současné době se ke stejnému účelu používá výpočetní technika, konkrétně **generátory pseudonáhodných čísel**. Tyto účelově vytvořené algoritmy jsou běžnou součástí softwaru různého typu a umožňují snadno realizovat náhodné výběry.

Pokud se můžeme spolehnout na to, že seznam jednotek v základním souboru je uspořádán náhodně, lze uplatnit tzv. **systematický výběr**. Nejprve stanovíme **výběrový krok** jako podíl $N/n = k$. Z prvních k jednotek v seznamu vybereme náhodně první jednotku systematického výběru a pořadové číslo každé další jednotky obdržíme opakovaným přičítáním kroku výběru k. Tak například při výběru pěti jednotek ze tří set je výběrový krok 60 (300/5 = 60). Z prvních šedesáti jednotek náhodně vybereme jednu jednotku, řekněme patnáctou. Dále pak vybereme jednotky s pořadovými čísly 75, 135, 195 a 255.

Předchozí výběrové postupy realizují stejné pravděpodobnosti vybrání jednotek v základním souboru. Pokud je chceme použít pro rozdílné pravděpodobnosti vybrání, pak zastoupíme jednotky tolika položkami v opoře výběru, aby bylo dosaženo odpovídajících rozdílů v pravděpodobnostech vybrání. Například dvojnásobné pravděpodobnosti vybrání je dosaženo tak, že příslušná jednotka je v seznamu uvedena dvakrát. Výhodnější je však samozřejmě opět použít počítačový algoritmus.

Neexistence opory výběru je samozřejmě závažnou (a velmi častou) komplikací. Je pak nutné zvážit jinou možnost výběru tak, aby nebyly narušeny jeho žádoucí vlastnosti. V terénu je možné realizovat systematický výběr například domácností formou tzv. náhodné procházky. Náhodně je stanovena startovací adresa (na níž se obvykle výzkum neprovádí) a dále pravidla pohybu v terénu (jakým směrem jít, jak směr měnit, jak vybírat byty v bytových domech atd.). Ve stanoveném směru jsou postupně odpočítávány domy či byty a vybrán je každý k-tý podobně jako u systematického výběru. Pravidla pohybu při náhodné procházce musí být formulována tak, aby zajišťovala skutečně pravděpodobnostní vybírání a znemožňovala jakákoliv subjektivní rozhodování o jejím průběhu.

V předchozím příkladu se v rámci procházky pohybujeme na území nějaké obce. Ale i tu jsme museli nejspíš vybrat, a možná i území, na kterém se nachází (např. okres). Výběr patrně proběhl ve více stupních. Často je totiž nutné z různých důvodů uvažovat o složitějším uspořádání pravděpodobnostního výběru. Důvody jsou dvojí: jednak organizační – především hledisko nákladů šetření, jednak kvalita úsudků o základním souboru, realizovaných následně. Vhodný způsob uspořádání vzorku totiž

může vést ke snížení velikosti chyb, kterých se dopouštíme, když o základním souboru činíme závěry pouze na základě výběrové informace.

Předpokládejme, že základní soubor lze rozdělit do menšího či většího počtu nepřekrývajících se skupin jednotek. Mohou přitom nastat dvě možnosti: při rozdělení souboru jsme se snažili o to, aby si jednotky ve skupinách byly co nejpodobnější a aby rozdíly z pohledu sledované veličiny byly spíše mezi skupinami, nebo jsou skupiny stejně různorodé, jako soubor sám.

V prvním případě provedeme z každé skupiny náhodný výběr vzorku jednotek. Hovoříme pak o **oblastním (stratifikovaném) výběru**. Oblastmi (straty) přitom nerozumíme pouze určitá území, ale také například věkové kategorie, domácnosti s určitým počtem členů, podniky tříděné podle počtu zaměstnanců a odvětví působnosti apod. Nejvýhodnější z hlediska následného zobecňování je situace, kdy rozsah výběru z každé oblasti je úměrný její velikosti (tedy z každé oblasti vybíráme stejné procento jednotek). V takovém případě je oblastní výběr dokonce účinnější a přináší kvalitnější informace o základním souboru než prostý náhodný výběr.

Ve druhém případě jsou skupiny pro takový postup příliš různorodé. Lépe bude náhodně vybrat skupiny a v jejich rámci pak provést vyčerpávající zjišťování. Hovoříme pak o **výběru skupin**. Pokud náhodně vybereme jednotky pouze v těchto vybraných skupinách, hovoříme o **dvoustupňovém výběru**. Uvedené postupy mohou být výhodné spíše organizačně než z hlediska kvality úsudků o základním souboru. V praxi pak je někdy nutno přistoupit k realizaci výběru i ve více než dvou stupních. Základní soubor domácností v České republice tvoří zhruba 4 miliony jednotek. Pokud má být realizováno jejich výběrové šetření, je vhodné rozdělit je podle krajů, dále podle okresů, podle sídel, u velkých sídel podle jejich částí. Příprava takového šetření, má-li přinést odpovídající výsledky, je z hlediska organizace náročná záležitost.

Nepravděpodobnostní výběry zákonitosti náhody v procesu vybírání pro úsudky o populaci využívat nemohou. Informace z nich získané se nedají objektivně zobecňovat, mohou mít tak pouze podpůrný a doplňující ráz. Značné přednosti nepravděpodobnostních výběrů však jsou v organizační oblasti, vyžadují nižší časové i finanční náklady než výběry pravděpodobnostní. Není tak divu, že v praxi jsou často upřednostňovány.

Bývají tak uplatňovány například příležitostné výběry (náhodné kontaktování například chodců na ulici) a samovolné výběry – ankety. Jejich velkým problémem je neznámá populace, z níž je výběr prováděn a na niž bychom mohli získané poznatky zobecnit. Mohou ji tvořit osoby s větším zájmem o téma šetření, ať už v pozitivním (obdivovatelé) či negativním smyslu (notoričtí stěžovatelé), osoby, které byly více než ostatní motivovány k účasti slibem nějaké odměny či výhry atd.

Úsudkové výběry jsou založeny na formulování určitého kritéria, jež má zajistit co nejlepší „logickou" shodu vzorku se základním souborem. Problém je ve způsobu, jakým je toto kritérium realizováno. Samotný akt volby jednotky probíhá na základě subjektivního rozhodnutí výběr provádějící osoby. Při použití **typického výběru** jsou

tak vybírány jednotky, považované v dané populaci za typické. V praxi terénních průzkumů je vůbec nejrozšířenějším výběrovým postupem **kvótní výběr.** Jeho cílem je zajistit shodnou strukturu základního a výběrového souboru vzhledem k několika podstatným (kvótním) znakům.

Kvótní znaky jsou jednoduché, snadno identifikovatelné a z hlediska základního souboru klíčové vlastnosti, jako je například věk, pohlaví, vzdělání nebo druh ekonomické aktivity. Podle hodnot kvótních znaků je základní soubor vlastně rozdělen do skupin a dodržení shodné struktury vzorku znamená, že počet vybíraných osob je úměrný velikosti skupiny. Za předpokladu dodržení náhodnosti volby jednotek by se jednalo o oblastní proporcionální výběr, který poskytuje relativně kvalitní úsudky o populaci. Samotná volba jednotky však náhodně neprobíhá, neboť o ní vedle kvótních znaků rozhoduje výběr realizující osoba (tazatel).

Námitku lze v případě kvótních výběrů vznést i proti myšlence dosažení shodné struktury vzhledem k několika kvótním znakům, jež nemusí nutně znamenat shodnou strukturu i z hlediska znaků ostatních, včetně zjišťovaných. Je pravda, že za kvótní jsou voleny takové znaky, s nimiž obvykle zjišťované znaky více či méně souvisejí. Na druhou stranu se praxe kvótních výběrů omezuje obvykle na dodržení pouze jednorozměrné, nikoliv vícerozměrné struktury základního souboru. Tak například vzorek se shoduje s populací vzhledem k podílu žen (například 55 %) a vzhledem k zastoupení jednotlivých věkových skupin (například nejmladší 30 %, středního věku 40 %, nejstarší 30 %). To ovšem neznamená, že se shoduje vzhledem k zastoupení žen v jednotlivých věkových skupinách. I přes zachování složení podle věku a pohlaví tak ve vzorku ve srovnání s populací mohou například převažovat mladší ženy a naopak starší muži.

Pro praxi kvótní výběry představují rychlejší, organizačně jednodušší a obvykle také levnější možnost vybírání než pravděpodobnostní postupy. Je proto pochopitelné, že jsou vůbec nejrozšířenější. Nutno říci, že jejich uživatelé vytvářením nejrůznějších kontrolních mechanismů usilují o co největší eliminování subjektivity rozhodování tazatelů, a tím o otupení nejčastějších výtek, jež jsou vůči kvótnímu výběru směřovány. Další argumenty (nejen) praxe ve prospěch kvótního výběru připomínají, že hodnocení kvality úsudků o cílové populaci je obvykle spojováno pouze s posouzením výběrové chyby a nepočítá s problémem **non-response**, tedy nezískáním informace od části vybraných jednotek.

Praxe výběrových statistických šetření osob se často musí vyrovnat s dost nízkou úspěšností při jejich kontaktování, ať už jsou její důvody jakékoliv. To ovšem může znamenat systematickou (nevýběrovou) chybu zjišťování, a tedy jeho významné zkreslení. Vyvstává pak pochopitelně otázka, jak s neúplnými údaji o některých jednotkách naložit, a zda si pravděpodobnostní výběry udrží své přednosti při vícerozměrném šetření s vysokým podílem chybějících hodnot.

S rozvojem internetu v posledních dvou desetiletích narůstá velmi rychle jeho využití ve všech fázích terénních šetření. Snižuje totiž významně náklady, přitom

umožňuje zkoumání relativně rozsáhlých souborů a dovoluje efektivní uplatnění nejrůznějších výzkumných nástrojů. Akademická sféra je (a měla by být) ve srovnání s praxí v rozvoji on-line výzkumů rezervovanější; hlavní příčinou jsou nároky na kvalitu dat, které internetové průzkumy často nesplňují a neumožňují tak realizovat objektivní závěry o zkoumaných populacích.

Nekvalifikované využití internetu může data i informace získané jejich analýzou významně ovlivnit. Z problematického definování populace, která je předmětem zkoumání, může vyplývat chyba jejího pokrytí v průběhu šetření. Platí to i pro země s relativně vysokým pokrytím internetem, jako je Česká republika, jak ukazuje konfrontace celkové a internetové populace země či celkové a internetové populace českých domácností. Zkoumaná populace proto bývá ve výzkumné praxi často zastoupena **on-line panelem**, z něhož jsou vzorky vybírány. To zase může vést k chybám v závislosti na tom, jakým způsobem je panel vytvořen. Způsob výběru z populace (nebo z on-line panelu) může vést až ke vzniku takového vzorku, který není reprezentativní a neumožňuje provést zobecnění na populaci buď vůbec, nebo jen velmi omezeným způsobem. I pravděpodobnostní výběr, jak jsme již konstatovali, může být problematický, je-li ovlivněn vyšší mírou non-response.

Předpokládejme nyní, že všechna úskalí spojená s výběrovým šetřením se podařilo uspokojivě překonat a máme ze vzorku k dispozici objektivní údaje. Statistická teorie pak umožňuje získané informace zobecnit na základní soubor, ale také získat představu o kvalitě takových úsudků. To vše ovšem pouze za předpokladu, že výběr byl realizován jako pravděpodobnostní. Nejjednodušší postupy zobecňování vycházejí **z prostého náhodného výběru** a navíc **s vracením**. Pokud nebude výslovně uvedeno jinak, budeme v této knize dále vycházet ze vzorků, pořízených popsaným způsobem.

Jak jsme již uvedli, v matematické statistice pracujeme nejen s daty pořízenými náhodným výběrem z konečné populace, ale obecněji s daty pořízenými náhodným výběrem z pravděpodobnostního rozdělení nějaké náhodné veličiny. Náhodný výběr o rozsahu n z rozdělení náhodné veličiny X je posloupnost n nezávislých, stejně rozdělených náhodných veličin s rozdělením stejným jako má náhodná veličina X. Můžeme si tak představit n nezávislých realizací (hodnot) náhodné veličiny X jako například n hodnot starobního důchodu v České republice v určitém měsíci, počet poruch v dílně v n vybraných dnech nebo informaci od n náhodně vybraných voličů o tom, zda by volili konkrétní stranu. Pokud budeme dále používat pojem náhodný výběr, budeme mít na mysli právě náhodný výběr z rozdělení náhodné veličiny.

3.2 Výběrová rozdělení

Zkoumanou náhodnou veličinu, kterou mohou být například měsíční příjem zaměstnance, volební preference nebo třeba měsíční spotřeba energie čtyřčlenné domácnosti, lze v základním souboru popsat pomocí různých popisných charakteristik. Pokud je soubor konečný, hovoříme o **populačních** (také **základních**) **charakteristikách**. Me-

zi nejdůležitější z nich patří u kvantitativních znaků průměr nebo rozptyl v základním souboru, u alternativních znaků relativní četnost nebo absolutní četnost v základním souboru. Tyto statistiky bychom byli teoreticky schopni nalézt provedením vyčerpávajícího šetření. Často se ale zabýváme základními soubory, které jsou tak velké, že již není účelné k nim přistupovat jako ke konečným, nebo zkoumáme proměnné, které se ke konečným základním souborům ani teoreticky nevztahují. Takové problémy zahrnují například úsudky vztahující se také na ještě nevyrobené výrobky, kdy základní soubor tvoří všechny výrobky vyráběné stejným postupem, tedy již vyrobené i ještě nevyrobené. Nebo chceme zkoumat spotřebu paliva na 100 km auta určitého typu, kdy je velmi těžké formulovat základní soubor tvořený všemi stokilometrovými úseky, které taková auta mohou ujet. V takových případech mluvíme o pravděpodobnostním rozdělení náhodné veličiny a o jeho charakteristikách jako je střední hodnota, rozptyl nebo pravděpodobnost nějakého sledovaného jevu. Někdy se zabýváme ne přímo charakteristikami, ale parametry pravděpodobnostních rozdělení. Například v normálním rozdělení můžeme zkoumat střední hodnotu. Uvážíme-li vlastnosti tohoto rozdělení, zajímáme se vlastně o parametr μ tohoto pravděpodobnostního rozdělení.

Hodnoty charakteristik rozdělení a parametrů nemůžeme obyčejně nijak přesně určit. V **klasickém**, někdy také **četnostním** přístupu předpokládáme, že tyto charakteristiky jsou neznámé konstanty. Tento přístup budeme používat v celém dalším textu. **Bayesovský přístup** (Bolstad a Curran, 2017) předpokládá, že tyto parametry jsou náhodné veličiny.

V dalším textu budeme náhodným výběrem rozumět náhodný výběr z rozdělení zkoumané proměnné. Z hodnot v náhodném výběru určíme výběrové protějšky k základním charakteristikám, těmto protějškům říkáme **výběrové charakteristiky** nebo také **statistiky**. Pojem statistika používáme tedy nejen jako název pro vědní oblast (ekonomická statistika, matematická statistika), ale také pro libovolné funkce náhodného výběru, jako je průměr, výběrový medián, variační koeficient a další. Charakteristiky základního souboru a parametry rozdělení sledované proměnné jsou pevné hodnoty. Statistiky jsou vypočítané z náhodného výběru, a tedy z náhodných veličin, jsou proto také náhodnými veličinami.

Nyní budeme uvažovat konečný základní soubor. Nejprve zavedeme potřebné značení, budeme využívat symboly a pojmy z kapitol o popisné statistice a teorii pravděpodobnosti. Náhodnou veličinu označíme X, počet statistických jednotek v základním souboru N a počet pozorování v náhodném výběru n. Hodnoty náhodné veličiny X pro jednotky v základním souboru budeme značit malými písmeny, tedy $x_1, x_2, \ldots, x_N$. Jednotky jsou vybrány náhodným výběrem, proto jsou pozorování náhodné veličiny a v souladu se značením z druhé kapitoly je budeme značit velkými písmeny $X_1, X_2, \ldots, X_n$. Jejich konkrétní hodnoty označíme malými písmeny $x_1, x_2, \ldots, x_n$.

Průměr v základním souboru (**základní** nebo také **populační průměr**) je

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{3.1}$$

jeho výběrovým protějškem je aritmetický průměr v náhodném výběru, tj.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{3.2}$$

Někdy nás bude zajímat součet hodnot proměnné X pro všechny jednotky v základním souboru (**základní** nebo **populační úhrn**) ve tvaru $\sum_{i=1}^{N} x_i = N\mu$.

Obdobně rozptyl v základním souboru (**základní** nebo **populační rozptyl**) je definován jako

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2, \tag{3.3}$$

jeho výběrový protějšek **výběrový rozptyl** je dán vztahy

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{n}{n-1}\left(\overline{X^2} - \bar{X}^2\right). \tag{3.4}$$

Výraz za druhým rovnítkem v (3.4) se využívá k výpočtu výběrového rozptylu a nazývá se také výpočetní tvar rozptylu. Všimněme si, že při výpočtu rozptylu v základním souboru postupujeme stejným způsobem jako při výpočtu rozptylu v první kapitole. Proto můžeme využít i všechny již uvedené vlastnosti. Koeficient $1/(n-1)$ v definici výběrového rozptylu (3.4) je volen tak, aby hodnota statistiky S^2 co nejlépe přibližovala neznámý základní rozptyl σ^2.

V první kapitole této knihy bylo ukázáno, že pro popis variability je užitečné používat charakteristiky, které mají stejné měrné jednotky jako zkoumaná proměnná, neboť v takovém případě jsou pozorované hodnoty i charakteristiky přímo srovnatelné. Proto budeme i v matematické statistice používat **směrodatnou odchylku** v základním souboru σ a **výběrovou směrodatnou odchylku** S, definované jako odmocniny z rozptylů (3.3) a (3.4), tedy

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \tag{3.5}$$

respektive

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}. \tag{3.6}$$

V případě alternativního znaku X sledujeme výskyt určité vlastnosti. Budeme předpokládat, že statistický znak nabývá hodnoty 1 v případě, že statistická jednotka má danou vlastnost, a 0 v opačném případě. Důležitými základními charakteristikami pak je počet jednotek s danou vlastností v základním souboru

$$m = \sum_{i=1}^{N} x_i \tag{3.7}$$

a relativní počet nositelů dané vlastnosti v základním souboru

$$\pi = \frac{m}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{3.8}$$

Výběrovými protějšky jsou **výběrový úhrn**

$$M = \sum_{i=1}^{n} X_i \tag{3.9}$$

a **výběrová relativní četnost**

$$P = \frac{M}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{3.10}$$

Příklad 3.1

Uvažujme firmu s 500 zaměstnanci, u kterých budeme sledovat měsíční mzdu v tis. Kč (náhodná veličina X), dobu práce ve firmě v měsících (náhodná veličina Y) a alternativní znak Z určující, zda zaměstnanec je žena. Platí tedy $N = 500$. Předpokládejme, že průměry v celé firmě jsou 31,5 tis. Kč a 45 měsíců, směrodatné odchylky jsou 12 tis. Kč a 23 měsíců. Jestliže ve firmě pracuje 130 žen, je $m = 130$ a $\pi = 130/500 = 0{,}26$ (26 % žen mezi zaměstnanci firmy). Každý zaměstnanec je jednou statistickou jednotkou, hodnoty sledovaných znaků pro j-tého zaměstnance ($j = 1, 2, \ldots, 500$) jsou x_j, y_j a z_j. V tomto případě by bylo snadné zjistit hodnoty populačních charakteristik pro všechny zaměstnance, tedy pro celou analyzovanou populaci o $N = 500$ jednotkách. Představme si ale, že provedeme náhodný výběr pěti zaměstnanců s vracením ($n = 5$) a pro tyto zaměstnance získáme hodnoty proměnných (pozorování) z tabulky 3.1.

Tab. 3.1 Výpočet výběrových charakteristik (statistik)

i	x_i	y_i	z_i	x_i^2	y_i^2
1	17,9	30	0	320,41	900
2	18,2	62	0	331,24	3 844
3	25,3	20	0	640,09	400
4	37,7	48	1	1 421,29	2 304
5	45,6	74	0	2 079,36	5 476
Součet	144,7	234	1	4 792,39	12 924
Průměr	28,9	46,8	0,2	958,48	2 584,8

Vybrali jsme $100(n/N) = 100(5/500) = 1$ % zaměstnanců. Pokud bychom uvažovali krok výběru, do výběru jsme zařadili každého stého zaměstnance, neboť je $N/n = 500/5 = 100$. Z posledního řádku tabulky 3.1 snadno dostaneme výběrové charakteristiky

$$\bar{x} = 28{,}9,\ \bar{y} = 46{,}8,\ p = 0{,}2, \tag{3.11}$$

$$s_X^2 = \frac{5}{4}(958{,}48 - 28{,}9^2) = 151{,}2,\ s_Y^2 = \frac{5}{4}(2\,584{,}8 - 46{,}8^2) = 493{,}2 \tag{3.12}$$

$$s_X = \sqrt{151{,}2} = 12{,}3,\ s_Y = \sqrt{493{,}2} = 22{,}2. \tag{3.13}$$

Při výpočtu charakteristik variability ve vzorci (3.12) byl použit výpočetní tvar pro rozptyl, uvedený v (3.4). Vypočtené hodnoty můžeme srovnat s charakteristikami v celé firmě (s populačními charakteristikami)

$$\mu_X = 31{,}5,\ \sigma_X = 12,\ \mu_Y = 45,\ \sigma_Y = 23,\ \pi = 0{,}26. \tag{3.14}$$

Pokud bychom vybrali jiné zaměstnance, výběrové charakteristiky (3.11) – (3.13) by se změnily, neboť závisejí na náhodném výběru. Základní charakteristiky pro celou firmu uvedené v (3.14) by zůstaly stejné, neboť závisejí pouze na základním souboru.

Předpokládejme ještě, že bylo dvakrát nezávisle vybráno sto zaměstnanců (každý pátý zaměstnanec, neboť 500/100 = 5) a byly zjištěny průměrné mzdy 32 tis. Kč a 28,9 tis. Kč a výběrové směrodatné odchylky 11,5 tis. Kč a 13 tis. Kč.

Připomeňme, že podle předchozí části 3.1 je náhodným výběrem z rozdělení mezd jen náhodný výběr s vracením, ve druhém případě jsou opakované výběry závislé, neboť není možné vybrat vícekrát jednoho pracovníka. Vzhledem k tomu, že výběry sta osob jsou 20krát větší než výběr pěti osob, aritmetický průměr ze sta hodnot by měl být přesnějším přiblížením průměru v celé firmě, než průměr, určený z hodnot pro pět zaměstnanců. Matematická statistika umožňuje tuto intuitivní úvahu kvantifikovat, neboli odpovědět na otázku, o kolik nebo kolikrát je aritmetický průměr přesnější jako přiblížení průměru v celé firmě. V dalším textu uvidíme, že odpověď dvacetkrát není správná, neboť je třeba uvážit i variabilitu mzdy ve firmě, popsanou rozptylem mezd.

■

Již jsme uvedli, že všechny výběrové charakteristiky uvedené výše jsou příkladem statistik. Pravděpodobnostní rozdělení používaná pro statistiky se nazývají výběrová rozdělení. Znalost pravděpodobnostního rozdělení statistik je základním nástrojem k tomu, abychom mohli zobecňovat hodnoty, získané z náhodného výběru na celý základní soubor. Například odpovědět na otázku, co nám říká průměrná mzda určená z pěti hodnot pro vybrané pracovníky o průměrné mzdě v celé firmě, nebo co lze usoudit z volebních preferencí tisíce vybraných voličů o názorech všech voličů v daném čase. Odpovědi na podobné otázky jsou tématem následujících částí, kdy použijeme metody popisné statistiky z první kapitoly pro hodnoty náhodného výběru a budeme usuzovat z náhodného výběru na celý základní soubor.

V matematické statistice existují dva základní typy induktivních úsudků, které usuzují z omezené informace, obsažené v náhodném výběru na celý základní soubor. Prvním jsou **odhady**, které na základě náhodného výběru přibližují hodnotu neznámé základní charakteristiky, druhým pak **testování hypotéz**. Již se nebudeme omezovat na konečné základní soubory a jejich základní charakteristiky. Pokud je základní soubor nekonečný, budeme se při zkoumání vlastností náhodné proměnné X zabývat její střední hodnotou $E(X)$ místo základním průměrem, rozptylem $D(X)$ místo základním rozptylem a pravděpodobností π místo základní relativní četností.

Navíc, pokud například řekneme, že výběrový průměr bude blízký (neznámému) základnímu průměru, doplníme také odpověď na otázku, jak blízké tyto dvě hodnoty

jsou. K tomu budeme potřebovat pojmy z teorie pravděpodobnosti, kterými se zabývá druhá kapitola této knihy. Ukážeme také, že je výhodné a dobře vysvětlitelné pro přiblížení hodnot neznámých charakteristik v základním souboru a parametrů pravděpodobnostních rozdělení konstruovat nějaký vhodný interval, který neznámé hodnoty bude s nějakou vysokou pravděpodobností obsahovat.

3.3 Bodový a intervalový odhad

Nejprve budeme neznámou hodnotu základní charakteristiky odhadovat jednou hodnotou, kterou nazveme **bodovým odhadem**. Takový odhad doplňujeme mírou jeho přesnosti, která zohledňuje kvalitu odhadu a je založena na velikosti kolísání daného odhadu v případě opakovaných výběrů. Druhou možností je neznámou charakteristiku odhadnout ne jedním číslem (bodem), ale intervalem, který s nějakou vysokou pravděpodobností neznámou hodnotu obsahuje. Takový odhad nazýváme **intervalovým odhadem**, sestrojený interval pak **intervalem spolehlivosti**. Nejprve se budeme zabývat bodovým odhadem charakteristik v základním souboru a parametrů pravděpodobnostních rozdělení.

Vzhledem k tomu, že hodnotu odhadované charakteristiky neznáme, nemůžeme kvalitu jejích bodových odhadů posuzovat prostým srovnáním charakteristiky a jejího odhadu, jako to bylo možné v příkladu 3.1. Proto se musíme spokojit s obecnějšími postupy, které budou založeny pouze na pravděpodobnostních rozděleních statistik.

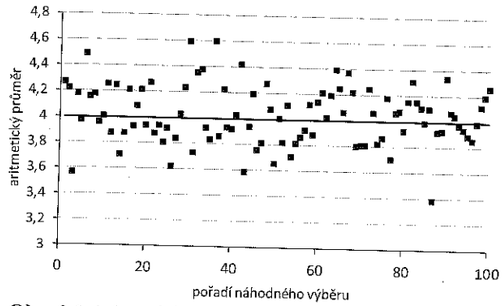
Předpokládejme, že nás bude zajímat průměrná měsíční mzda v České republice v určitém měsíci. Populace osob, které mzdu pobírají, je konečná a teoreticky by bylo možné průměrnou mzdu, tedy základní průměr, určit úplným šetřením. Můžeme tedy hovořit o populačním průměru mzdy a také o střední mzdě, oba termíny se vztahují ke stejné hodnotě. Pokud by nás ovšem zajímala střední hodnota spotřeby benzínu na sto kilometrů určitého typu auta, i při daném typu a vybavení nemůžeme hledanou hodnotu nalézt. Neexistuje totiž žádný konečný statistický soubor stokilometrových jízd, dokonce ani soubor všech aut, uvažujeme-li například ještě auta nevyrobená a jízdy ještě neuskutečněné. V tomto případě můžeme hovořit jen o střední spotřebě benzínu.

Předpokládejme, že máme k dispozici náhodný výběr $X_1, X_2, \ldots, X_n$ z rozdělení náhodné veličiny X, tedy informaci o mzdě n vybraných příjemců mzdy nebo o spotřebě benzínu na n náhodně vybraných stokilometrových jízd auta. Neznámou střední hodnotu náhodné veličiny $E(X)$ nebo průměr μ v základním souboru odhadneme aritmetickým průměrem $\bar{X}$, určeným podle vztahu (3.2). Statistiku $\bar{X}$ nazveme **bodovým odhadem**, někdy také používáme název **estimátor**. Dále již budeme hovořit jen o střední hodnotě, všechna tvrzení ovšem platí i pro průměr v konečném základním souboru. Již víme, že aritmetický průměr je **výběrovým protějškem** střední hodnoty. Střední hodnota je konstanta, o které předpokládáme, že je neznámá. Aritmetický průměr hodnot v náhodném výběru je náhodná veličina, neboť jeho hodnota závisí na konkrétních pozorováních v náhodném výběru. Můžeme ovšem určit střední hodnotu

aritmetického průměru, která na hodnotách náhodného výběru již nezávisí. Použijeme-li vlastnosti střední hodnoty probrané ve druhé kapitole této knihy a stejné rozdělení všech veličin v náhodném výběru, dostáváme

$$E(\bar{X}) = E(X).$$

Střední hodnota $\bar{X}$ je tedy rovna odhadované charakteristice bez ohledu na to, že neví-me, kolik její hodnota je. Při opakovaných náhodných výběrech tedy hodnota aritmetického průměru kolísá kolem střední hodnoty rozdělení, jak je ukázáno na obrázku 3.1. Bylo generováno 100 náhodných výběrů o rozsahu $n = 20$ z normálního rozdělení se střední hodnotou 4 a směrodatnou odchylkou 1. Na obrázku 3.1 jsou znázorněny aritmetické průměry určené z jednotlivých náhodných výběrů a je patrné, že skutečně kolísají kolem střední hodnoty rozdělení 4. Průměrná hodnota vypočtených průměrů na obrázku je 4,032 a směrodatná odchylka je rovna 0,224.



Obr. 3.1 Aritmetický průměr v opakovaných náhodných výběrech

Odhady, které mají tuto vlastnost, nazýváme **nezkreslenými** (též **nestrannými** nebo **nevychýlenými**). Takové odhady neznámou hodnotu charakteristiky systematicky ani nenadhodnocují ani nepodhodnocují, jejich hodnoty při opakování náhodného výběru kolísají kolem odhadované hodnoty tak, jak jsme viděli na obrázku 3.1.

Velikost tohoto kolísání popisuje chyba odhadu, čím odhad kolísá více, tím je chyba větší, a tedy odhad je méně přesný. Ve druhé kapitole byla velikost kolísání hodnot náhodné veličiny popisována rozptylem nebo směrodatnou odchylkou, tyto charakteristiky tedy využijeme i v tomto případě. Pro aritmetický průměr dostáváme

$$D(\bar{X}) = \frac{D(X)}{n}. \tag{3.15}$$

Ze vztahu vyplývá, že směrodatná odchylka aritmetického průměru jako bodového odhadu střední hodnoty, tj.

$$\sqrt{D(\bar{X})} = \sqrt{\frac{D(X)}{n}} \tag{3.16}$$

roste s rozptylem $D(X)$ sledované náhodné veličiny X a klesá s rozsahem výběru. Tento vztah odpovídá intuitivní představě, že čím větší je náhodný výběr, tím více informace máme o střední hodnotě a tím přesnější tedy bude její bodový odhad. Míra, uvedená v (3.16), se nazývá **směrodatná chyba odhadu**. Bodové odhady doplňujeme touto hodnotou, neboť teprve potom je bodový odhad dobře interpretovatelný.

V této chvíli již můžeme odpovědět na otázku, kterou jsme si položili na konci příkladu 3.1. Směrodatná odchylka aritmetického průměru pro 5 osob je $12/\sqrt{5} = 5{,}4$ a pro 100 osob je $12/\sqrt{100} = 1{,}2$. Směrodatná odchylka prvního odhadu je přibližně 4,5 krát větší než směrodatná odchylka druhého, odhad sestrojený na základě výběru 100 osob je tedy jen 4,5 krát přesnější, i když máme k dispozici dvacetkrát více pozorování. Směrodatná odchylka aritmetických průměrů znázorněných na obrázku 3.1 je rovna $1/\sqrt{20} = 0{,}223\,6$ a je téměř totožná s již uvedenou výběrovou hodnotou 0,223 8.

Pokud chceme odhadnout základní úhrn, nezkresleným odhadem je veličina $N\bar{X}$. Směrodatná chyba tohoto odhadu je

$$\sqrt{D(N\bar{X})} = N\sqrt{\frac{D(X)}{n}}. \tag{3.17}$$

Tento postup můžeme použít také pro libovolný počet jednotek N, kdy budeme uvažovat například celkovou střední mzdu 10 pracovníků ($N = 10$) nebo celkovou střední spotřebu auta na 1500 km (15 stokilometrových jízd, $N = 15$).

Pokud je odhad zkreslený, nejsou rozptyl nebo směrodatná odchylka odhadu vhodnou mírou přesnosti odhadu, neboť nás zajímá kolísání kolem odhadované charakteristiky, nikoliv kolem střední hodnoty odhadu. V takovém případě používáme **střední čtvercovou chybu odhadu,** která je v případě nezkresleného odhadu rovna rozptylu odhadu. Pro odhady, které jsou zkreslené, je rozptyl zvýšen o penalizaci, která zohledňuje vzdálenost střední hodnoty odhadu a odhadované charakteristiky.

Další důležitou vlastností, kterou od bodových odhadů očekáváme, je **konzistence odhadu**. Řekneme, že odhad je konzistentní, jestliže se zvyšujícím se rozsahem výběru se odhad přibližuje (konverguje) k odhadované hodnotě. Požadavek odráží představu, že s větším počtem pozorování získáváme více informace o neznámé odhadované charakteristice, a odhad by tedy měl být blíže odhadované hodnotě. Konvergence odhadů jako náhodných veličin, používaná v tomto případě, se nazývá konvergence podle pravděpodobnosti. S touto konvergencí jsme se seznámili ve druhé kapitole, v části věnované limitním větám. Aritmetický průměr je konzistentním odhadem střední hodnoty pro libovolné rozdělení, které má konečný rozptyl a platí

$$\bar{X} \to E(X) \text{ podle pravděpodobnosti.}$$

Pokud si vybíráme jeden odhad z několika nezkreslených, konzistentních odhadů, volíme obyčejně ten, který má nejmenší rozptyl, a je tedy z uvažovaných odhadů nejpřesnější. Lze si snadno představit, že pokud máme pevný rozsah náhodného výběru, nelze sestrojit odhad s libovolně malou směrodatnou odchylkou (podrobnosti jsou uvedeny v Malá, 2013). Pro konstrukci odhadu totiž můžeme využít pouze informaci obsaženou v datech, ta je ale pro pevný rozsah náhodného výběru daná a omezená. Odhad, jehož rozptyl dosahuje minimálního možného rozptylu a je tedy nejpřesnější z těch, které z daných dat můžeme sestrojit, nazveme **vydatným**. Již jsme uvedli, že aritmetický průměr je nezkresleným a konzistentním odhadem střední hodnoty. Pokud má navíc zkoumaná náhodná veličina X normální rozdělení, je i vydatným odhadem.

Ve vzorcích (3.15) a (3.17) jsme k nalezení přesnosti odhadu potřebovali znát rozptyl rozdělení $D(X)$. Obyčejně ale rozptyl náhodné veličiny X neznáme, proto je třeba ho také odhadnout. Rozptyl rozdělení náhodné veličiny X bodově odhadujeme **výběrovým rozptylem** S^2, definovaným v (3.4). Tento odhad je nezkresleným odhadem rozptylu, a tedy platí

$$E(S^2) = D(X). \tag{3.18}$$

Tato vlastnost je důvodem pro použití koeficientu $1/(n-1)$ v definici výběrového rozptylu místo přirozenější volby $1/n$. V případě větších výběrů není velký rozdíl mezi použitím hodnoty $1/(n-1)$ a $1/n$, pro malé výběry, ale rozdíl může být patrný. Pro $n = 5$ z příkladu 3.1 dostáváme koeficienty 0,25 a 0,2. Odhad S^2 je také konzistentní. Pokud bychom chtěli odhadnout směrodatnou odchylku $\sqrt{D(X)}$, použijeme výběrovou směrodatnou odchylku $S = \sqrt{S^2}$. Tento odhad již není nezkreslený, jeho střední hodnota ale konverguje k odhadované charakteristice s rostoucím rozsahem výběru (pro n jdoucí k nekonečnu, neboli pro $n \to \infty$). Takovým odhadům říkáme **asymptoticky nezkreslené**. Nyní můžeme odhadnout také směrodatnou chybu aritmetického průměru z (3.16) jako $S/\sqrt{n}$ a směrodatnou chybu odhadu populačního úhrnu z (3.17) jako $NS/\sqrt{n}$.

Nyní ukážeme jinou možnost, jak odhadnout střední hodnotu. Budeme předpokládat, že zkoumaná náhodná veličina má normální rozdělení, později předpoklad rozdělení odstraníme v případě velkého rozsahu náhodného výběru. Místo bodového odhadu a jeho přesnosti sestrojíme interval, který bude neznámou hodnotu obsahovat s předem danou vysokou pravděpodobností. Budeme hovořit o intervalovém odhadu střední hodnoty nebo také o intervalu spolehlivosti pro střední hodnotu. Vzhledem k tomu, že máme k dispozici jen omezenou informaci o odhadované charakteristice obsaženou v náhodném výběru, je možné, že sestrojený interval odhadovanou hodnotu neobsahuje.

Zvolíme pravděpodobnost $1 - \alpha$ pro nějaké $0 < \alpha < 1$, kterou nazveme **koeficient spolehlivosti**, někdy též jednoduše **spolehlivost** intervalového odhadu, a budeme hovořit o $100(1-\alpha)\%$ intervalu spolehlivosti. Je zřejmé, že máme snahu konstruovat intervalové odhady s velkou spolehlivostí, a tedy volit $1 - \alpha$ blízké jedné. Na druhé straně uvidíme, že odhady s velkou spolehlivostí mohou být velmi široké a mohou mít tedy jen omezené použití. Například interval 0,1 až 1 toho velmi málo říká o neznámé pravděpodobnosti, stejně jako interval (0; 91) o střední hodnotě měsíčních mezd

v tisících korun ve firmě. Jako kompromis mezi spolehlivostí a šířkou intervalů spolehlivosti volíme hodnotu α mezi 0,1 a 0,01. Používáme tak 90% až 99% intervaly spolehlivosti, nejčastěji se setkáme s 95% intervalem spolehlivosti. Meze intervalů jsou statistiky určené z náhodného výběru, jedná se tedy o náhodné veličiny a výsledné intervaly nazveme náhodnými intervaly. V matematické statistice používáme buď **oboustranné intervaly spolehlivosti**, ve kterých hledáme dolní i horní mez intervalu spolehlivosti, nebo intervaly **jednostranné**, kdy nás zajímá buď jen horní mez (hovoříme o **pravostranném intervalu spolehlivosti**), nebo jen dolní mez pro odhadovanou charakteristiku. V takovém případě se jedná o **levostranný interval spolehlivosti**.

Postup konstrukce a další pojmy spojené s intervaly spolehlivosti ukážeme na příkladu intervalu spolehlivosti pro neznámou střední hodnotu $E(X)$ normálního rozdělení se známým rozptylem. Pro normální rozdělení s parametry μ a σ^2 platí $E(X) = \mu$, a tedy se současně jedná o konstrukci intervalového odhadu pro tento parametr. Již víme, že bodovým odhadem μ je průměr $\bar{X}$ (píšeme $\hat{\mu} = \bar{X}$) a směrodatná chyba tohoto odhadu je podle (3.16) rovna $\sigma/\sqrt{n}$. Pak oboustranným $100(1 - \alpha)\%$ intervalem spolehlivosti pro parametr μ normálního rozdělení budeme nazývat interval (μ_D, μ_H) s náhodnými mezemi μ_D a μ_H. Tyto meze jsou konstruovány na základě náhodného výběru tak, aby platilo, že interval s danou spolehlivostí obsahuje hledanou charakteristiku, tedy

$$P(\mu_D < \mu < \mu_H) = 1 - \alpha. \tag{3.19}$$

Takových intervalů bychom ale mohli sestrojit velké množství. Budeme tedy navíc předpokládat, že interval (μ_D, μ_H) byl sestrojen tak, že platí

$$P(\mu < \mu_D) = \alpha/2 \quad \text{a} \quad P(\mu_H < \mu) = \alpha/2. \tag{3.20}$$

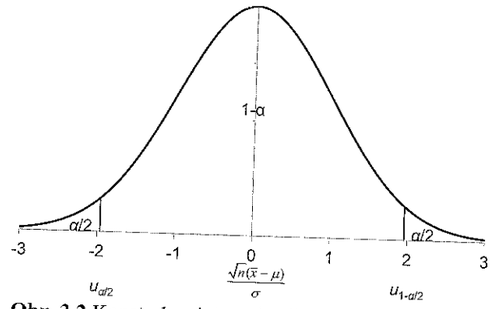
Pokud je skutečná střední hodnota menší než μ_D, sestrojený interval spolehlivosti střední hodnotu přecenil. Pokud je skutečná střední hodnota větší než μ_H, interval spolehlivosti střední hodnotu podcenil. Pravděpodobnosti takových chybných rozhodnutí jsou podle (3.20) rovny $\alpha/2$.

Intervaly spolehlivosti konstruujeme pomocí bodového odhadu (ve sledovaném případě $\bar{X}$) a dále nějaké statistiky se známým pravděpodobnostním rozdělením, obsahující tento odhad a hledaný parametr. Touto statistikou je

$$U = \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma}, \tag{3.21}$$

která má normované normální rozdělení $N(0, 1)$. Podle definice kvantilů můžeme psát (na obrázku 3.2 je zvoleno $\alpha = 0{,}05$, $u_{0{,}025} = -1{,}96$ a $u_{0{,}975} = 1{,}96$)

$$P\left(u_{\alpha/2} < \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} < u_{1-\alpha/2}\right) = 1 - \alpha. \tag{3.22}$$



Obr. 3.2 Konstrukce intervalového odhadu střední hodnoty

Ze symetrie normálního rozdělení plyne, že $u_{\alpha/2} = u_{1-\alpha/2}$. Proto můžeme (3.22) přepsat jako

$$P\left(-u_{1-\alpha/2} < \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} < u_{1-\alpha/2}\right) = 1 - \alpha.$$

Nyní budeme upravovat nerovnice uvnitř pravděpodobnosti tak, abychom dostali tvar (3.19). Po jednoduchých algebraických úpravách dostaneme

$$P\left(\bar{X} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Hledaným intervalovým odhadem (μ_D, μ_H) střední hodnoty μ je tedy interval

$$\left(\bar{X} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right). \tag{3.23}$$

Takto budeme v dalším textu intervaly spolehlivosti zapisovat. Krajní body intervalu jsou opravdu náhodné veličiny, které závisejí na aritmetickém průměru ve výběru $\bar{X}$, který je náhodnou veličinou na známých hodnotách σ, n a zvolené hodnotě α prostřednictvím kvantilu $u_{1-\alpha/2}$. Interval (3.23) lze také zapsat ve tvaru

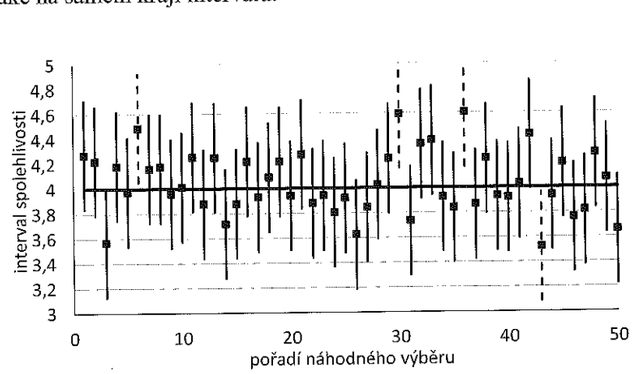
$$\bar{X} \pm u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = \bar{X} \pm \Delta = \left(\bar{X} - \Delta,\ \bar{X} + \Delta\right), \tag{3.24}$$

kdy bodový odhad $\bar{X}$ střední hodnoty μ je středem intervalu spolehlivosti. Šířka intervalu spolehlivosti je $2u_{1-\alpha/2}\sigma/\sqrt{n} = 2\Delta$ a pro pevné hodnoty σ, n a $u_{1-\alpha/2}$ je konstantní. Poloviční šířku intervalu spolehlivosti jsme označili Δ.

Všimněme si, že šířka oboustranného intervalu spolehlivosti závisí na směrodatné chybě odhadu střední hodnoty $\sigma/\sqrt{n}$. Čím je interval spolehlivosti širší, tím poskytuje méně informace o odhadované střední hodnotě, a tedy je méně přesný. Proto se šířce

intervalových odhadů budeme v dalším textu věnovat. Uvidíme, že velmi spolehlivé intervaly s koeficientem spolehlivosti blízkým jedné budou méně přesné, neboť budou příliš široké a naopak. Již zmíněné běžně používané hodnoty koeficientu spolehlivosti v intervalu 0,9–0,99 (pro $0{,}01 \le \alpha \le 0{,}1$) by měly přibližně zajišťovat rovnováhu mezi spolehlivostí a šířkou intervalového odhadu.

Uveďme ještě četnostní význam intervalu spolehlivosti, který nám pomůže správně intervaly spolehlivosti interpretovat. Představme si, že sestrojíme intervaly spolehlivosti pro určitou charakteristiku na základě velkého počtu nezávislých výběrů o rozsahu n z daného rozdělení. Potom bude přibližně $100(1-\alpha)\%$ takových intervalů obsahovat skutečnou hodnotu charakteristiky a zbytek tuto hodnotu obsahovat nebude. Na obrázku 3.3 jsou znázorněny 95% intervaly spolehlivosti pro $\mu = 4$ v rozdělení $N(4,1)$ sestrojené pro prvních 50 náhodných výběrů použitých na obrázku 3.1. Čárkovaně jsou znázorněny intervaly spolehlivosti, které odhadovanou hodnotu nepřekrývají. V tomto případě jsou to 4 z 50 (tři intervaly přeceňují, jeden podceňuje hodnotu 4), což je 8 % sestrojených intervalů. S rostoucím počtem opakování náhodného výběru by se toto procento přibližovalo hodnotě 5 %. Všimněme si, že všechny intervaly jsou stejně široké ($2u_{0,975}/\sqrt{20} = 0{,}877$), vypočtené aritmetické průměry jsou ve středu intervalů a pokud interval spolehlivosti hodnotu čtyři obsahuje, může být tato hodnota také na samém kraji intervalu.



Obr. 3.3 Opakované generované nezávislé výběry

Levostranným intervalem spolehlivosti pro parametr μ budeme rozumět interval s náhodnou mezí μ_D ve tvaru (μ_D, ∞), pro který platí

$$P(\mu_D < \mu) = 1 - \alpha. \tag{3.25}$$

Statistika μ_D je $100(1-\alpha)\%$ **dolní mez** pro μ. Odvodit hodnotu μ_D lze obdobně jako v případě oboustranného intervalu. V tomto případě ale není možné odhadovanou hodnotu přecenit, proto pravděpodobnost α není třeba rozdělit na dvě části. Z nerovnice

$$P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < u_{1-\alpha}\right) = 1-\alpha \tag{3.26}$$

dostaneme algebraickými úpravami uvnitř závorky

$$P\left(\bar{X} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}} < \mu\right) = 1-\alpha. \tag{3.27}$$

Dolní mezí pro μ je tedy $\bar{X} - u_{1-\alpha}\sigma/\sqrt{n}$ a levostranný interval spolehlivosti pro tento parametr můžeme zapsat jako

$$\left(\bar{X} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right). \tag{3.28}$$

Obdobně $100(1-\alpha)\%$ pravostranný interval spolehlivosti pro parametr μ s vlastností $P(\mu < \mu_H) = 1 - \alpha$ má tvar

$$\left(-\infty, \bar{X} + u_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right), \tag{3.29}$$

a $100(1-\alpha)\%$ **horní mez** pro μ je rovna $\bar{X} + u_{1-\alpha}\sigma/\sqrt{n}$.

Příklad 3.2

Test studijních předpokladů je konstruován tak, aby výsledky studentů měly normální rozdělení. Předpokládejme dále, že ze zkušenosti je známo, že variabilita výsledků studentů v testu, popsaná směrodatnou odchylkou, je 15 bodů. Pro 25 náhodně vybraných studentů středních odborných škol byl určen průměrný výsledek 77 bodů. Najdeme 95% oboustranný interval spolehlivosti pro střední výsledek testu mezi studenty odborných středních škol.

Řešení

Použijeme-li vzorec (3.23), dostaneme meze hledaného intervalu (μ_D, μ_H) ve tvaru $(u_{1\text{-}0{,}05/2} = u_{0,975} = 1{,}96)$

$$\left(\bar{x} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \left(77 - 1{,}96\cdot\frac{15}{\sqrt{25}},\ 77 + 1{,}96\cdot\frac{15}{\sqrt{25}}\right) = (71{,}12;\ 82{,}88).$$

Bodovým odhadem neznámé střední hodnoty je 77 bodů, $15/\sqrt{25} = 3$ body je jeho směrodatná chyba. Sestrojený interval neznámou hodnotu obsahuje s pravděpodobností 0,95, nebo přesněji s pravděpodobností 0,95 překrývá skutečný střední počet bodů, pokud interval položíme na množinu všech možných hodnot μ (0; 100). Spo-

lehlivost (95%) sestrojeného intervalu si můžeme představit tak, že v případě velkého počtu opakování výběru 25 studentů v průměru 95 % sestrojených intervalů bude obsahovat neznámou hodnotu a 5 % tuto hodnotu obsahovat nebude.

■

Nyní se opět zabývejme vztahem mezi spolehlivostí a přesností intervalového odhadu. Již víme, že spolehlivost odhadu zpravidla volíme. Uvažujeme-li oboustranný interval podle (3.24), je jeho šířka, která charakterizuje přesnost intervalového odhadu, rovna

$$2\Delta = 2\frac{u_{1-\alpha/2}\sigma}{\sqrt{n}}. \tag{3.30}$$

Vzhledem k tomu, že parametr σ je pevný, snadno vidíme, že pro zvolený koeficient spolehlivosti šířka intervalu klesá s rostoucím rozsahem výběru, neboť $\sqrt{n}$ je ve jmenovateli zlomku. Na druhou stranu pro daný rozsah výběru n šířka intervalu roste (a přesnost klesá) s rostoucím koeficientem spolehlivosti a jeho šířka klesá (a přesnost roste), zmenšuje-li se koeficient spolehlivosti. Zvolíme-li pro pevný rozsah výběru příliš vysoký koeficient spolehlivosti, interval může být velmi široký, a tedy nepřesný.

Pokud je hodnota Δ dána před provedením náhodného výběru, pak udává požadovanou přesnost budoucího odhadu a nazývá se **přípustná chyba**. Známe-li hodnotu rozptylu σ^2 a máme předepsanou přípustnou chybu, můžeme určit potřebný počet pozorování k dosažení dané přesnosti intervalu pro zvolený koeficient spolehlivosti intervalového odhadu. Z rovnice (3.30) dostaneme nerovnici

$$\Delta \geq \frac{u_{1-\alpha/2}\sigma}{\sqrt{n}},$$

jejímž řešením je minimální potřebný rozsah výběru n ve tvaru

$$n \geq \left(\frac{u_{1-\alpha/2}\sigma}{\Delta}\right)^2. \tag{3.31}$$

Z nerovnice (3.31) je zřejmé, že čím je menší přípustná chyba, tím více pozorování je třeba získat pro dosažení této přesnosti odhadu při zvolené spolehlivosti. Je tedy možné interval spolehlivosti udělat libovolně přesný, pro takový případ ale může být potřebný počet pozorování nedosažitelný. Při praktickém použití tedy vždy musíme mít na paměti nejen požadovanou přesnost budoucích úsudků, ale také finanční, technické, časové nebo etické limity získávání náhodných výběrů.

Příklad 3.3

Je známo, že střední doba životnosti výrobku je 5000 hodin a směrodatná odchylka je rovna 200 hodin; budeme předpokládat přibližně normální rozdělení této doby. Předpokládejme, že změna výrobního postupu prodlouží životnost, ale variabilitu neovlivní. Najdeme počet výrobků, jejichž životnost je třeba prošetřit, aby přípustná chyba 95 % intervalu spolehlivosti pro střední dobu životnosti byla maximálně 50 hodin. Jinak řečeno najdeme rozsah náhodného výběru potřebný k tomu, aby šířka intervalu spolehlivosti byla maximálně 100 hodin.

Řešení

Podle předpokladu známe variabilitu $\sigma^2 = 200^2$. Ze vzorce (3.31) dostaneme minimální potřebný rozsah náhodného výběru ve tvaru

$$n \geq \left(\frac{u_{1-\alpha/2}\sigma}{\Delta}\right)^2 = \left(\frac{1{,}96 \cdot 200}{50}\right)^2 = 61{,}5,$$

po zaokrouhlení (nahoru) by bylo třeba mít aspoň 62 pozorování k dosažení předepsané přesnosti intervalového odhadu.

■

V předchozím textu byly sestrojeny intervaly spolehlivosti pro střední hodnotu normálního rozdělení za předpokladu, že je znám rozptyl rozdělení. Taková situace je ale výjimečná a nejčastěji hledáme intervaly spolehlivosti v případě, kdy rozptyl normálního rozdělení znám není. V takovém případě můžeme směrodatnou odchylku rozdělení, kterou potřebujeme pro konstrukci intervalového odhadu, odhadnout výběrovou směrodatnou odchylkou S. Pro konstrukci intervalů spolehlivosti pak používáme kvantily t_P Studentova rozdělení s $n - 1$ stupni volnosti.

Obdobným způsobem jako výše získáme intervaly spolehlivosti (Malá, 2013)

$$\left(\bar{X} - t_{1-\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right), \tag{3.32}$$

$$\left(-\infty,\ \bar{X} + t_{1-\alpha}\frac{S}{\sqrt{n}}\right) \text{ a } \left(\bar{X} - t_{1-\alpha}\frac{S}{\sqrt{n}},\ \infty\right).$$

S rostoucím rozsahem náhodného výběru se kvantily Studentova rozdělení přibližují kvantilům normovaného normálního rozdělení. Pokud používáme pouze tabulky Studentova rozdělení, od 30 stupňů volnosti nahradíme kvantil Studentova rozdělení kvantilem normovaného normálního rozdělení. V případě použití softwaru se doporučuje používat kvantily Studentova rozdělení i pro náhodné výběry o více než třiceti pozorováních.

Sestrojený interval má opět ve svém středu hodnotu bodového odhadu $\bar{X}$ a je typu $\bar{X} \pm \Delta$. Jeho šířka 2Δ je rovna

$$2t_{1-\alpha/2}\frac{S}{\sqrt{n}}.$$

Na rozdíl od intervalu spolehlivosti v případě známého rozptylu je náhodnou veličinou, protože závisí na náhodné veličině S.

Příklad 3.4

Uvažujme zařízení, které zhotovuje balíčky cukru. Budeme předpokládat, že hmotnost balíčku má normální rozdělení. Při kontrole množství obsahu bylo náhodně vybráno a zváženo 20 balíčků a byla zjištěna průměrná hmotnost 0,977 kg a výběrová směrodatná odchylka 0,015 kg. Najdeme 95% oboustranný interval spolehlivosti pro střední hmotnost balíčku a dále pro tuto charakteristiku najdeme 95% dolní mez. Totéž uděláme pro případ, ve kterém stejné statistiky byly určeny z hmotností 80 náhodně vybraných balíčků.

Řešení

Podle (3.32) dostáváme oboustranný interval ($t_{0,975}(19) = 2,093$)

$$\left(0,977 - 2,093 \cdot \frac{0,015}{\sqrt{20}},\ 0,977 + 2,093 \cdot \frac{0,015}{\sqrt{20}}\right) = (0,970; 0,984),$$

a dolní mez ($t_{0,95}(19) = 1,729$)

$$0,977 - 1,729 \cdot \frac{0,015}{\sqrt{20}} = 0,971\,2.$$

Oboustranný interval spolehlivosti pro střední hodnotu balíčku (v kilogramech) je tedy (0,970; 0,984) a se spolehlivostí 95 % tento interval obsahuje neznámou střední hmotnost balíčku. Na základě nalezené dolní meze můžeme říci, že se spolehlivostí 95 % je neznámá střední hmotnost balíčku větší než hodnota 0,971 kg.

Ještě najdeme intervaly spolehlivosti pro $n = 80$. V tomto případě již použijeme kvantil normovaného normálního rozdělení $u_{0,975} = 1,96$ a dostáváme interval spolehlivosti pro střední hodnotu ve tvaru

$$\left(0,977 - 1,96 \cdot \frac{0,015}{\sqrt{80}},\ 0,977 + 1,96 \cdot \frac{0,015}{\sqrt{80}}\right) = (0,974; 0,980),$$

a dolní mez ($u_{0,95} = 1,645$)

$$0,977 - 1,645 \cdot \frac{0,015}{\sqrt{80}} = 0,974\,2.$$

Všimněme si, že pro větší počet pozorování je oboustranný interval spolehlivosti pro střední hmotnost užší a dolní mez je větší, obě hodnoty jsou bližší bodovému odhadu střední hmotnosti balíčku 977 gramů. Uvedli jsme, že pro normální rozdělení je střední hodnota $E(X)$ rovna parametru μ normálního rozdělení. Tato rovnost platí pro číselné hodnoty; rozdíl je v tom, že střední hmotnost balíčku udáváme v kilogramech nebo gramech, a všechny parametry pravděpodobnostních rozdělení, tedy i μ, jsou bezrozměrná čísla. ■

Nyní sestrojíme interval spolehlivosti pro rozptyl normálního rozdělení. Tento interval již není typu (3.24), protože pro jeho konstrukci nemůžeme využít symetrická rozdělení jako jsou normální nebo Studentovo rozdělení. Vycházíme ze statistiky $[(n-1)S^2]/\sigma^2$, která má rozdělení χ^2 s $n-1$ stupni volnosti (viz druhá kapitola). Použijeme kvantil χ^2_P tohoto rozdělení a oboustranný interval spolehlivosti pro rozptyl rozdělení má pak tvar

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}},\ \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right). \tag{3.33}$$

Na rozdíl od intervalů pro parametr μ tento interval již není symetrický kolem bodového odhadu rozptylu S^2, nicméně tuto hodnotu vždy obsahuje. Navíc je třeba v tabulkách nebo v softwaru najít dva kvantily χ^2 rozdělení – jeden dolní kvantil $\chi^2_{\alpha/2}(n-1)$ a jeden horní kvantil $\chi^2_{1-\alpha/2}(n-1)$. Slova dolní a horní se vztahují k velikosti pravděpodobnosti, pro kterou kvantil hledáme.

Často nás místo rozptylu σ^2 zajímá směrodatná odchylka σ. Z nerovnosti

$$\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}}$$

můžeme po odmocnění dostat oboustranný interval spolehlivosti pro σ ve tvaru

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}},\ \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}}}\right). \tag{3.34}$$

Ještě sestrojíme jednostranné intervaly pro rozptyl normálního rozdělení. Všimněme si, že jednostranné intervaly můžeme dostat tak, že použijeme odpovídající mez oboustranného intervalu, musíme jen změnit pravděpodobnost v kvantilu – použijeme tedy kvantily $\chi^2_{1-\alpha}(n-1)$ a $\chi^2_{\alpha}(n-1)$. Potom pro rozptyl rozdělení σ^2 dostaneme $100(1-\alpha)\%$ dolní mez pro rozptyl rozdělení ve tvaru

$$\frac{(n-1)S^2}{\chi^2_{1-\alpha}}$$

a obdobně $100(1-\alpha)\%$ horní mez

$$\frac{(n-1)S^2}{\chi^2_{\alpha}}.$$

Uvážíme-li, že rozptyl nemůže být záporný, vyjádříme jednostranné intervaly ve tvaru

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha}}, \infty\right) \text{ a } \left(0, \frac{(n-1)S^2}{\chi^2_{\alpha}}\right).$$

Pro směrodatnou odchylku σ jsou dolní a horní mez po řadě

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha}}} \text{ a } \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha}}}.$$

Odpovídající jednostranné intervaly spolehlivosti pro směrodatnou odchylku mají tvar

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha}}}, \infty\right) \text{ a } \left(0, \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha}}}\right).$$

Příklad 3.5

Vraťme se k příkladu 3.4. Sestrojíme 95% interval spolehlivosti pro směrodatnou odchylku hmotnosti balíčku.

Řešení

Podle zadání příkladu je $n = 20$ a výběrová směrodatná odchylka je rovna 0,015 kg, a tedy $s^2 = 0{,}015^2 = 0{,}000\,225$. Použijeme-li dále vzorec (3.34), je $n - 1 = 20 - 1 = 19$, $\alpha = 0{,}05$, $\chi^2_{0{,}025}(19) = 8{,}907$ a $\chi^2_{0{,}975}(19) = 32{,}852$. Potom hledaný interval spolehlivosti dostaneme ve tvaru

$$\left(\sqrt{\frac{(20-1)\cdot 0{,}000\,225}{32{,}852}}, \sqrt{\frac{(20-1)\cdot 0{,}000\,225}{8{,}907}}\right) = (0{,}012;\ 0{,}022).$$

Všimněme si, že bodový odhad směrodatné odchylky 0,015 kg leží uvnitř sestrojeného intervalu spolehlivosti, není ale v jeho středu, je blíže dolní než horní hranici intervalu.

■

Do této chvíle jsme předpokládali, že sledovaná náhodná veličina X má normální rozdělení. Často se ale setkáváme s úlohou sestrojit interval spolehlivosti pro střední hodnotu $E(X)$ nějakého obecného rozdělení.

Z druhé kapitoly této knihy již víme, že pro velký rozsah výběru platí centrální limitní věta Lindebergova-Lévyho a aritmetický průměr $\bar{X}$ pozorování v náhodném výběru má přibližně rozdělení $N(E(X), D(X)/n)$. Aritmetický průměr je dále konzistentním a nezkresleným odhadem střední hodnoty. Pro směrodatnou odchylku odhadu platí (3.16) a pokud rozptyl náhodné veličiny neznáme, použijeme bodový odhad $S/\sqrt{n}$.

Předpokládejme tedy, že pravděpodobnostní rozdělení náhodné veličiny X má konečný rozptyl a rozsah výběru je dostatečně velký, abychom mohli použít centrální limitní větu. Pro velké rozsahy výběru přibližně platí

$$P\left(\bar{X} - u_{1-\alpha/2}\sqrt{\frac{S^2}{n}} < E(X) < \bar{X} + u_{1-\alpha/2}\sqrt{\frac{S^2}{n}}\right) = 1 - \alpha.$$

Oboustranný (asymptotický) interval spolehlivosti pro $E(X)$ dostaneme ve tvaru

$$\left(\bar{X} - u_{1-\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X} + u_{1-\alpha/2}\frac{S}{\sqrt{n}}\right).$$

Jednostranné intervaly mají tvar

$$\left(\bar{X} - u_{1-\alpha}\frac{S}{\sqrt{n}}, \infty\right) \text{ a } \left(-\infty, \bar{X} + u_{1-\alpha}\frac{S}{\sqrt{n}}\right).$$

Příklad 3.6

Pro 50 náhodně vybraných pracovníků velké firmy byla zjištěna průměrná mzda 29 500 Kč a směrodatná odchylka 2500 Kč. Odhadneme bodově střední mzdu ve firmě, najdeme 95% oboustranný interval spolehlivosti pro střední mzdu a dále 95% dolní mez pro střední mzdu.

Řešení

Bodovým odhadem střední mzdy je statistika $\bar{x} = 29\,500$ Kč, jeho směrodatnou chybu odhadneme jako

$$\frac{s}{\sqrt{n}} = \frac{2500}{\sqrt{50}} = 354 \text{ Kč}.$$

Pro 95% interval spolehlivosti je $\alpha = 0{,}05$ Proto kvantily, které použijeme pro konstrukci intervalu spolehlivosti, jsou $u_{1-\alpha/2} = u_{0{,}975} = 1{,}96$ (pro oboustranný interval) a $u_{1-\alpha} = u_{0{,}95} = 1{,}645$ (pro jednostranný interval). Potom dostáváme

$$\left(\bar{x} - u_{1-\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + u_{1-\alpha/2}\frac{s}{\sqrt{n}}\right) = (29\,500 - 1{,}96\cdot 354,\ \ 29\,500 + 1{,}96\cdot 354),$$

tedy 28 806 až 30 194 korun.

Hledáme 95% dolní mez pro střední mzdu ve firmě

$$\bar{x} - u_{1-\alpha}\frac{s}{\sqrt{n}} = 29\,500 - 1{,}645\cdot 354 = 28\,918 \text{ (Kč)}.$$

■

Nyní se budeme zabývat odhadem pravděpodobnosti nějakého jevu, například pravděpodobnosti, že výrobek má určitou vadu nebo že volič bude volit konkrétní politickou stranu. Předpokládejme, že máme k dispozici pozorování, která vznikla nezávislým opakováním příslušného náhodného pokusu, proto někdy mluvíme o úsudcích o parametru π alternativního rozdělení. Zjistíme tedy u náhodně vybraných výrobků, zda mají nebo nemají vadu, nebo u náhodně vybraných voličů, zda budou volit danou stranu. Bodovým odhadem pravděpodobnosti π je (výběrová) relativní četnost P výskytu zkoumaného jevu v posloupnosti n nezávislých náhodných pokusů. Budeme psát

$$\hat{\pi} = P.$$

Z výsledků druhé kapitoly plyne, že pro statistiku P platí

$$E(P)=\pi \text{ a } D(P)=\frac{\pi(1-\pi)}{n}. \tag{3.35}$$

Odhad P je nezkresleným, konzistentním a vydatným odhadem pravděpodobnosti π a směrodatná chyba odhadu je rovna

$$\sqrt{D(P)}=\sqrt{\frac{\pi(1-\pi)}{n}}.$$

Vzhledem k tomu, že pravděpodobnost π neznáme, k odhadu směrodatné chyby použijeme statistiku

$$\sqrt{\frac{P(1-P)}{n}}. \tag{3.36}$$

Ještě doplníme intervaly spolehlivosti pro pravděpodobnost π. Budeme předpokládat, že rozsah výběru je dostatečně velký na to, abychom mohli použít centrální limitní větu Moivreovu-Laplaceovu (Marek, 2012), a hledaný asymptotický oboustranný interval spolehlivosti (π_D, π_H) zapíšeme ve tvaru

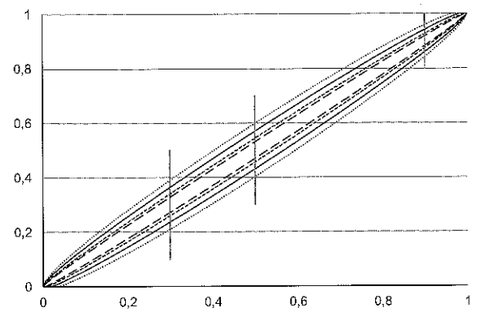
$$\left(P-u_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}, P+u_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}\right), \tag{3.37}$$

nebo také jako

$$P\pm u_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}. \tag{3.38}$$

Parametr π alternativního rozdělení může nabývat pouze hodnot z intervalu (0;1), proto jednostranné intervaly budou mít tvar

$$\left(0, P+u_{1-\alpha}\sqrt{\frac{P(1-P)}{n}}\right) \text{ a } \left(P-u_{1-\alpha}\sqrt{\frac{P(1-P)}{n}}, 1\right).$$



Obr. 3.4 Intervaly spolehlivosti pro parametr π

Na obrázku 3.4 je pro rozsahy výběrů 100, 200, 500 a 1000 znázorněna závislost šířky oboustranného intervalu spolehlivosti na rozsahu výběru a hodnotě bodového odhadu P. Nejširší interval spolehlivosti je při pevném rozsahu výběru pro $P = 0{,}5$ a jeho šířka klesá se vzdáleností odhadu P od hodnoty 0,5 stejně na obě strany. Víme, že interval spolehlivosti se zkracuje s rozsahem výběru a s klesající spolehlivostí.

Často je cílem analýzy odhadnout absolutní četnost jednotek se zkoumanou vlastností v základním souboru $N\pi$, kde N je rozsah základní populace. Můžeme se například ptát na počet voličů konkrétní strany místo na pravděpodobnost, že volič bude volit tuto stranu, nebo na počet domácností vychovávajících aspoň jedno dítě v daném regionu místo pravděpodobnosti, že ve vybrané domácnosti budou vyrůstat děti. Počet jednotek základní populace N obyčejně známe, proto populační úhrn $N\pi$ můžeme odhadnout statistikou NP. Všimněme si, že hodnota této statistiky nemusí být (a téměř nikdy není) celé číslo. Z vlastností rozptylu dostáváme také směrodatnou chybu tohoto odhadu ve tvaru (společně s jejím odhadem, neboť parametr π neznáme)

$$N\sqrt{\frac{\pi(1-\pi)}{n}}, \text{ resp. } N\sqrt{\frac{P(1-P)}{n}}.$$

Tento postup často používáme i pro jiný počet jednotek N, než je rozsah základního souboru (například chceme odhadnout střední počet výrobků, které koupí $N = 20$ potenciálních zákazníků na základě šetření $n = 399$ zákazníků). Odhadovanému počtu událostí pak říkáme úhrn. Z náhodného výběru máme informaci o n statistických jednotkách, zajímá nás však situace pro N jiných (nepozorovaných) jednotek.

Oboustranný interval spolehlivosti v takovém případě má tvar (podle (3.37))

$$(N\pi_D, N\pi_H)=\left(NP-Nu_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}, NP+Nu_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}\right),$$

jednostranné intervaly pak

$$\left(0, NP+Nu_{1-\alpha}\sqrt{\frac{P(1-P)}{n}}\right) \text{ a } \left(NP-Nu_{1-\alpha}\sqrt{\frac{P(1-P)}{n}}, N\right).$$

Příklad 3.7

Vedení firmy, která má 3000 zaměstnanců, uvažuje o zavedení nového benefitu. Bylo vybráno 200 zaměstnanců, z nichž zájem projevilo 80. Najdeme 95% oboustranný interval spolehlivosti pro pravděpodobnost, že zaměstnanec bude mít o benefit zájem, a také pro střední počet zaměstnanců firmy, kteří by měli o benefit zájem.

Řešení

Bodovým odhadem pravděpodobnosti je relativní četnost $p = 80/200 = 0{,}4$, směrodatná chyba odhadu je rovna $\sqrt{(0{,}4\cdot 0{,}6)/200} = 0{,}034\,6$. Bodovým odhadem počtu zá-

jemců ve firmě je 40 procent z 3000, tedy 3000·0,4 = 120 osob. Směrodatná chyba úhrnu je 3000·0,034 6 = 103,9 osob. Pro interval spolehlivosti dostáváme

$$p \pm u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0{,}4 \pm 1{,}96\sqrt{\frac{0{,}4 \cdot 0{,}6}{200}} = 0{,}4 \pm 0{,}067\,9,$$

tedy neznámá pravděpodobnost je (se zvolenou spolehlivostí 95 %) prvkem intervalu (0,332 1; 0,467 9). Střední počet zájemců o benefit je se spolehlivostí 0,95 prvkem intervalu

(3000 · 0,332 1; 3000 · 0,467 9) = (996,3; 1403,7),

po zaokrouhlení na přirozená čísla dostaneme interval spolehlivosti pro střední počet zájemců (996; 1404). Všimněme si, že při násobení bodového odhadu tisíci je třeba uvažovat bodový odhad s dostatečnou přesností, proto byla zvolena čtyři desetinná místa.

■

3.4 Testování statistických hypotéz

Nyní se budeme věnovat druhému základnímu postupu statistické indukce, a to testování statistických hypotéz. Nejprve uvedeme základní východiska a zavedeme potřebné pojmy, které jsou společné pro všechny testy, se kterými se můžeme ve statistické literatuře setkat. Testování hypotéz je všeobecně používaným postupem v analýze dat a dostupnost software umožňuje testovat hypotézy širokému spektru uživatelů. Interpretace výsledků statistického testu ovšem není úplně snadná, a proto bude úvodu do testování hypotéz věnována část 3.4.1. Dále představíme některé vybrané statistické testy.

3.4.1 Základy testování hypotéz

Při statistickém testování hypotéz pracujeme s nějakou hypotézou, která je formulována obvykle nejprve slovně jako výrok, teprve pak ji přeformulujeme do matematických symbolů, například do testu hypotézy o parametru nebo parametrech nějakého pravděpodobnostního rozdělení. **Statistické hypotézy** odrážejí řešený konkrétní problém a vycházejí z **výzkumných hypotéz**. Výsledky statistického testování by měly sloužit jako kvantitativní podklad pro rozhodování o výzkumných hypotézách, je proto třeba dobře rozlišovat případnou **významnost statistickou** (plynoucí z testu statistické hypotézy) a **významnost skutečnou** nebo **faktickou** (plynoucí z povahy řešeného problému). Rozhodnutí o výzkumné hypotéze je vždy třeba učinit po zvážení všech znalostí a informací, nejen výsledků statistického testu.

Při testování hypotéz může jít o hypotézu, že o nový produkt bude mít zájem aspoň 20 % potenciálních zákazníků, nebo že střední doba do poruchy určitého výrobku bude aspoň 3 roky. Pokud tyto hypotézy neplatí, bude zájem menší než 20 % zákazníků a doba životnosti bude v průměru kratší než 3 roky. Platnost statistické hypotézy budeme posuzovat na základě náhodného výběru a budeme se ptát, jestli pozorování odporují nebo neodporují zvolené hypotéze. V případě hypotézy o zájmu o určitý výrobek zřejmě 30 zájemců ze 100 dotázaných hovoří pro platnost hypotézy, 10 zájemců ze 100 dotázaných proti hypotéze. Ve druhém příkladu zřejmě průměrná doba do poruchy 5 let, nalezená v náhodném výběru, hovoří pro platnost hypotézy, průměrná doba 2 roky proti hypotéze. Při testování hypotéz je třeba stanovit, které hodnoty statistik určené z náhodného výběru hovoří pro hypotézu a které proti ní. V našich příkladech bychom měli problém rozhodnout při 21 zájemcích o produkt ze 100 dotázaných nebo třeba průměrné doby do poruchy 2,9 roku. V případě, že hodnoty v náhodném výběru hypotéze odporují, hypotézu zamítneme, v opačném případě ji nezamítneme a usoudíme, že pozorované hodnoty hypotéze neodporují. Všimněme si již v této chvíli, že hypotézu (budeme ji nazývat nulová hypotéza nebo testovaná hypotéza) nepotvrzujeme, tvrzení je slabší. Říkáme, že data hypotéze neodporují nebo také že nejsou v rozporu s hypotézou. Při testu nulové hypotézy předpokládáme, že platí, dokud neusoudíme, že náhodný výběr takovému předpokladu neodpovídá. Proti nulové hypotéze stavíme alternativní hypotézu, pro kterou se rozhodneme v případě zamítnutí nulové hypotézy. Obdobně pomocí testu hypotézy nelze žádné tvrzení dokázat nebo prokázat, takto silná tvrzení statistický test neumožňuje.

Postup testování hypotéz ukážeme na výše zmíněném příkladu. Odvodíme test, který se nazývá test o parametru alternativního rozdělení π. V našem případě nás zajímá pravděpodobnost, že zákazník bude mít zájem o nový produkt. Předpokládejme, že vedení firmy se rozhodlo uvést na trh nový produkt jen tehdy, pokud o něj bude mít zájem více než 20 % potenciálních zákazníků. Zvolíme nulovou hypotézu H_0 ve tvaru

$$H_0: \pi = 0{,}2 \tag{3.39}$$

a **alternativní hypotézu** H_1 (někdy jen **alternativu**) jako

$$H_1: \pi > 0{,}2. \tag{3.40}$$

Pokud by vedení firmy zajímalo, jestli o produkt bude mít zájem 20 % zákazníků, zvolíme alternativu, že jich nebude 20 %, tedy

$$H_1: \pi \neq 0{,}2. \tag{3.41}$$

Alternativní hypotéza (3.40) se nazývá **jednostranná**, hypotéza (3.41) je **oboustranná**. Vhodnou hypotézu a alternativu volíme podle řešeného problému; všimněme si, že znaménko rovnosti patří vždy do nulové hypotézy.

Při konstrukci testu zvolíme **testové kritérium** (někdy také říkáme **testovou statistiku**), jehož hodnotu počítáme z náhodného výběru. Všechny dosažitelné hodnoty testové statistiky rozdělíme na dvě disjunktní části, **obor přijetí** V a **kritický obor** W. V tomto textu budeme předpokládat, že sjednocení obou množin V a W vyčerpává všechny hodnoty, kterých může nabývat testové kritérium. Obecně tomu ale tak být nemusí, v některých případech například test pro určité hodnoty **mlčí** (neumí rozhodnout), jak je tomu například u Durbinova-Watsonova testu v páté kapitole. Obor přijetí V obsahuje takové hodnoty testové statistiky, které nejsou v rozporu s platností nulové hypotézy a pro které nulovou hypotézu nezamítáme. Kritický obor W obsahuje hodnoty testové statistiky, které jsou v rozporu s nulovou hypotézou (hovoří tedy pro alter-

nativní hypotézu) a vedou k zamítnutí nulové hypotézy. Pokud došlo k zamítnutí nulové hypotézy, nazveme test **statisticky významným**. V opačném případě řekneme, že test je **nevýznamný**.

Vzhledem k tomu, že se rozhodujeme na základě náhodného výběru, a tedy jen s omezenou informací o základním souboru, může i správně zvolený a použitý test hypotézy vést k mylnému rozhodnutí. V tabulce 3.2 jsou znázorněna správná a chybná rozhodnutí, spojená s testováním hypotéz. V řádcích tabulky je uvedena (neznámá) skutečnost a ve sloupcích pak rozhodnutí na základě testu hypotézy. Na hlavní diagonále jsou správná rozhodnutí, v prvním řádku správné rozpoznání platné nulové hypotézy a ve druhém řádku správné rozpoznání platné alternativní hypotézy. Na vedlejší diagonále v tabulce jsou špatná rozhodnutí. Nesprávné zamítnutí platné nulové hypotézy (zamítneme nulovou hypotézu, i když ve skutečnosti platí) nazýváme **chyba prvního druhu**. V tomto případě také hovoříme o falešném zamítnutí nulové hypotézy. Jestliže naopak nulovou hypotézu nezamítneme, i když ve skutečnosti neplatí (nesprávné nezamítnutí nulové hypotézy), dopouštíme se **chyby druhého druhu**. Zda jsme se některé chyby dopustili, nevíme, matematická statistika ale umožňuje určit aspoň pravděpodobnosti takových chybných rozhodnutí.

Tab. 3.2 Správná a chybná rozhodnutí při testování hypotéz

skutečnost / rozhodnutí	nezamítáme H_0	zamítáme H_0
H_0	správné rozhodnutí	chyba 1. druhu
H_1	chyba 2. druhu	správné rozhodnutí

Jistě by bylo dobré, aby obě pravděpodobnosti obou výše zmíněných chybných rozhodnutí byly malé. Bohužel platí, že snižujeme-li pravděpodobnost jednoho špatného rozhodnutí, zvyšuje se pravděpodobnost druhého špatného rozhodnutí. Budeme-li konstruovat test tak, aby byla velmi malá pravděpodobnost, že zamítneme nulovou hypotézu, i když ve skutečnosti platí, zřejmě se nám může častěji stát, že při této opatrnosti nezamítneme nulovou hypotézu, i když bychom již měli. Chybná rozhodnutí testu vždy mají následky, proto je třeba jim věnovat pozornost. V předchozí části jsme si vybírali odhady, které byly co nejpřesnější. V případě testů hypotéz chceme mít testy takové, že pro pevnou pravděpodobnost chyby prvního druhu mají co nejmenší pravděpodobnost chyby druhého druhu. Doplňková pravděpodobnost chyby druhého druhu se nazývá **síla testu** a snažíme se konstruovat a používat testy, které mají co největší sílu. Znamená to, že při volbě pravděpodobnosti chyby prvního druhu test co nejlépe poznává platnost alternativní hypotézy.

Pravděpodobnost chyby prvního druhu obyčejně volíme stejně jako jsme volili spolehlivost u intervalových odhadů, a značíme ji α. Tuto pravděpodobnost nazýváme také **hladina významnosti** nebo **hladina testu**. Pravděpodobnost chyby druhého druhu, která odpovídá zvolenému testu, zvolené hladině významnosti a náhodnému výběru, který máme k dispozici, je nutno vypočítat nebo aspoň odhadnout. Zkoumáním pravděpodobnosti chyby druhého druhu se potom zabývá **analýza síly testu**.

Jak jsme již objasnili, rozhodnutí o obou hypotézách v teorii statistického testování hypotéz nejsou stejně silná. Podle předchozího textu volíme jedno z následujících rozhodnutí:

- nezamítáme nulovou hypotézu na hladině významnosti α,
- zamítáme nulovou hypotézu na hladině významnosti α, a to ve prospěch alternativní hypotézy.

Nulovou hypotézu tedy nepřijímáme, ale pouze konstatujeme, že ji na základě náhodného výběru nemůžeme zamítnout. Můžeme také říci, že data neodporují nulové hypotéze. I když tedy vše vztahujeme k nulové hypotéze, rozhodnutí o alternativní hypotéze je silnější, neboť nulovou hypotézu zamítáme ve prospěch platnosti alternativní hypotézy (tedy že platí alternativa).

Ke konstrukci kritického oboru využíváme známé pravděpodobnostní rozdělení testového kritéria a hledáme kvantily tohoto rozdělení tak, aby pravděpodobnost chyby prvního druhu byla rovna zvolenému α. Tyto kvantily vymezují kritický obor, proto je nazýváme **kritickými hodnotami**. Některé statistické tabulky obsahují přímo kritické hodnoty, které se pak hledají podle zvolené hladiny významnosti. Zde ale budeme i nadále při konstrukci a zápisu kritických oborů používat pouze kvantily. Kritický obor W, pro který je pravděpodobnost chyby prvního druhu α, budeme značit W_α, odpovídající obor přijetí V_α. Vyznačíme tak hladinu významnosti, na které obě množiny závisejí.

Ve statistickém softwaru je nejčastěji pro rozhodnutí používána ***p*-hodnota**. Tuto p-hodnotu určíme jako pravděpodobnost, že za platnosti nulové hypotézy obdržíme hodnotu testového kritéria získanou z náhodného výběru, nebo hodnotu ještě méně příznivou nulové hypotéze. Je to také minimální hladina významnosti, pro kterou ještě zamítáme nulovou hypotézu. Proto také místo pojmu p-hodnota někdy používáme pojem **minimální hladina významnosti** nebo **dosažená hladina významnosti**. Postup, jak p-hodnotu vypočítat, ukážeme dále u konkrétních testů. P-hodnotu můžeme používat k rozhodnutí o nulové hypotéze, nikoliv například na srovnávání výsledků různých testů nebo statistických modelů. Při rozhodnutí o hypotéze porovnáváme p-hodnotu se zvolenou hladinou významnosti α, pokud platí:

- p-hodnota $\leq \alpha$, zamítneme nulovou hypotézu ve prospěch alternativy,
- p-hodnota $> \alpha$, nulovou hypotézu nezamítneme.

Nízká hodnota tedy hovoří pro alternativní hypotézu, vysoká hodnota pro nulovou hypotézu. Mezí, dělící obě rozhodnutí, je zvolená hladina významnosti α.

Vraťme se k testu o parametru alternativního rozdělení s nulovou hypotézou (3.39) a jednou z alternativ (3.40) a (3.41). Problém lze obecně zapsat jako test nulové hypotézy

$$H_0: \pi = \pi_0 \tag{3.42}$$

proti oboustranné hypotéze

$$H_1: \pi \neq \pi_0, \tag{3.43}$$

nebo proti některé ze dvou jednostranných hypotéz

$$H_1: \pi < \pi_0 \quad \text{a} \quad H_1: \pi > \pi_0. \tag{3.44}$$

Předpokládejme, že máme k dispozici náhodný výběr o rozsahu n. Pokud náhodný výběr obsahuje pouze malé množství pozorování, je třeba použít některý test, určený pro malé náhodné výběry. V tomto textu budeme dále předpokládat, že náhodný výběr je dostatečně velký na to, aby platila centrální limitní věta. V takovém případě můžeme test založit na testovém kritériu

$$U = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{\sqrt{n}(P-\pi_0)}{\sqrt{\pi_0(1-\pi_0)}}, \tag{3.45}$$

které má za platnosti nulové hypotézy (3.42) asymptoticky normované normální rozdělení. Podle tabulky 3.3 závisí kritický obor na zvolené alternativní hypotéze. Je-li alternativní hypotéza oboustranná, hovoří proti nulové hypotéze hodnoty testového kritéria buď velmi malé, nebo velké. Hodnoty blízké nule svědčí pro nulovou hypotézu, neboť platí-li nulová hypotéza, je relativní četnost P přibližně rovna pravděpodobnosti π, a tedy čitatel zlomku je malý. Pro jednostranné alternativní hypotézy hovoří vždy jen příliš malé, nebo jen příliš velké hodnoty testového kritéria.

Tab. 3.3 Test o parametru π ve velkém náhodném výběru

H_0	H_1	Testové kritérium	Kritický obor
$\pi = \pi_0$	$\pi < \pi_0$	$U = \frac{\sqrt{n}(P-\pi_0)}{\sqrt{\pi_0(1-\pi_0)}} \quad U \approx N(0,1)$	$W_\alpha = \{u;\ u \le -u_{1-\alpha}\}$
	$\pi > \pi_0$		$W_\alpha = \{u;\ u \ge u_{1-\alpha}\}$
	$\pi \neq \pi_0$		$W_\alpha = \{u;\ \lvert u\rvert \ge u_{1-\alpha/2}\}$

P-hodnotu pro jednotlivé alternativy v (3.43) a (3.44) lze určit jako

- $\pi < \pi_0$: p-hodnota $= P(U \le u) = \Phi(u)$,
- $\pi > \pi_0$: p-hodnota $= P(U \ge u) = 1 - \Phi(u)$,
- $\pi \neq \pi_0$: p-hodnota $= 2P(U \ge |u|) = 2(1 - \Phi(|u|))$.

Všimněme si, že součet p-hodnot pro obě jednostranné alternativy je roven jedné a menší z obou p-hodnot je rovna polovině p-hodnoty pro oboustrannou alternativu. Proto známe-li p-hodnotu pro jednu z alternativ, umíme dopočítat p-hodnoty i pro další dvě alternativy. Této vlastnosti využívají některé implementace v softwaru. Je ale patrné, že určení správné p-hodnoty, která odpovídá konkrétní použité alternativní hypotéze, je třeba věnovat pozornost.

Vraťme se nyní k problému popsanému hypotézou (3.39) a jednostrannou alternativou (3.40). Zvolíme hladinu významnosti $\alpha = 0{,}05$. Chyba prvního druhu znamená, že test doporučí dát na trh produkt, který nebude dostatečně úspěšný. Pravděpodobnost této chyby je tedy 0,05. Chyba druhého druhu v tomto případě znamená, že nepoznáme platnost alternativy, a tedy firma nedá na trh produkt, který by mohl být úspěšný. Předpokládejme, že ze 160 dotázaných respondentů by zvážilo koupi nového produktu 44 osob, což je 100(44/160) = 27,5 % dotázaných. Máme tedy posoudit, zda toto procento, určené z náhodného výběru, je již (statisticky) významně větší než 20 % v nulové hypotéze. Z (3.45) po dosazení $\pi_0 = 0{,}2$, $n = 160$ a $p = 0{,}275$ dostaneme

$$u = \frac{\sqrt{n}(p-\pi_0)}{\sqrt{\pi_0(1-\pi_0)}} = \frac{\sqrt{160}(0{,}275-0{,}20)}{\sqrt{0{,}2 \cdot 0{,}8}} = 2{,}372.$$

Zvolili jsme hladinu významnosti 0,05 a podle tabulky 3.3 je

$$W_{0,05} = \{u;\ u \ge u_{1-0,05}\} = \langle u_{0,95}, \infty) = \langle 1{,}645;\ \infty).$$

Hodnota $u = 2{,}372$ testového kritéria je v kritickém oboru, a tedy na základě náhodného výběru zamítneme hypotézu, že zájem má 20 % zaměstnanců, ve prospěch alternativy, že zájem je větší. Stejný závěr učiníme pomocí p-hodnoty. Z kritického oboru plyne, že hodnoty testové statistiky ještě méně příznivé nulové hypotéze než hodnota 2,372 jsou prvkem intervalu $\langle 2{,}372, \infty)$. Za platnosti nulové hypotézy má testové kritérium normované normální rozdělení, proto je třeba určit pravděpodobnost

$$P(U \ge 2{,}172) = 1 - \Phi(2{,}372) = 0{,}009.$$

p-hodnota je tedy rovna 0,009 a dostaneme stejné rozhodnutí, neboť

$$0{,}009 = p\text{-hodnota} < \alpha = 0{,}05.$$

V případě oboustranné alternativy (3.41) stačí jen nalézt nový kritický obor. Podle tabulky 3.3 dostáváme

$$W_{0,05} = \{u;\ |u| \ge u_{1-0,05/2}\} = \left(-\infty, -u_{1-0,05/2}\right\rangle \cup \left\langle u_{1-0,05/2},\ \infty\right) = \left(-\infty; -1{,}96\right\rangle \cup \left\langle 1{,}96; \infty\right)$$

a platí $u = 2{,}372 \in W_{0,05}$. Proto i tuto alternativu zamítneme. Stejný výsledek bychom získali pomocí p-hodnoty, neboť z výsledků pro jednostrannou alternativu a úvahy za definicí p-hodnoty dostáváme

$$p\text{-hodnota} = 2 \cdot 0{,}009 = 0{,}018 < 0{,}05.$$

V literatuře nebo ve výsledcích testů implementovaných v softwaru se někdy setkáváme také s prezentací výsledků testů pomocí hvězdiček, více hvězdiček označuje menší p-hodnotu. Nejčastější postup v matematické statistice je volit

- * pro $0{,}01 \le p\text{-hodnota} < 0{,}05$,
- ** pro $0{,}001 \le p\text{-hodnota} < 0{,}01$,
- *** $p\text{-hodnota} < 0{,}001$

Je-li test nevýznamný a nulová hypotéza není zamítnuta, používá se zkratka NS (**N**on **S**ignificant). V předchozím případě bychom jednostranný test označili dvěma hvězdičkami (**) a oboustranný test jednou hvězdičkou (*).

Předchozí úvahy shrneme do několika kroků, které je třeba při testování hypotéz vždy udělat. Kroky 1 až 4 jsou velmi důležité a i v případě použití výpočetní techniky je třeba je důsledně provést. U implementovaných programů je třeba navíc znát postup, jak data zadat a jak správně přečíst a interpretovat výsledky.

1. Formulace problému statistického testu na základě výzkumné hypotézy, nalezení vhodného testu, volba nulové a alternativní hypotézy, volba hladiny významnosti testu.
2. Výpočet hodnoty testového kritéria, stanovení pravděpodobnostního rozdělení testového kritéria za platnosti nulové hypotézy.
3. Nalezení kritické hodnoty (kritických hodnot), sestrojení kritického oboru nebo určení p-hodnoty.
4. Rozhodnutí o výsledku testu a interpretace výsledků.

V dalším textu podrobně probereme některé nejčastěji používané testy.

3.4.2 Testy o parametrech na základě jednoho výběru

Předpokládejme nyní, že náhodný výběr pochází z normálního rozdělení $N(\mu,\sigma^2)$ a budeme testovat hypotézu o střední hodnotě μ. Nulová hypotéza má tvar

$$H_0: \mu = \mu_0, \tag{3.46}$$

kde μ_0 je hypotetická střední hodnota. Podle řešeného problému můžeme volit buď oboustrannou alternativu ve tvaru

$$H_1: \mu \neq \mu_0, \tag{3.47}$$

nebo některou z jednostranných alternativních hypotéz

$$H_1: \mu < \mu_0 \quad \text{a} \quad H_1: \mu > \mu_0. \tag{3.48}$$

Nejprve je třeba, stejně jako v případě konstrukce intervalu spolehlivosti, rozhodnout, zda je nebo není známa hodnota rozptylu σ^2. Předpokládejme nejprve, že rozptyl rozdělení je znám. V takovém případě použijeme testovou statistiku

$$U = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n},$$

která má za platnosti nulové hypotézy (3.46) normované normální rozdělení. Pokud rozptyl znám není, použijeme stejnou statistiku jako při konstrukci intervalu spolehlivosti pro střední hodnotu

$$T = \frac{\bar{X} - \mu_0}{S}\sqrt{n}, \tag{3.49}$$

která má (podle druhé kapitoly) za platnosti nulové hypotézy rozdělení $t(n-1)$. Proto se tento test nazývá **jednovýběrový *t*-test**. Pokud platí nulová hypotéza, budou hodnoty testové statistiky blízké nule, neboť v takovém případě bude rozdíl mezi $\bar{X}$ a μ_0 malý. Hodnoty testové statistiky T vzdálené od nuly na obě strany (pro oboustrannou alternativu $\mu \neq \mu_0$) nebo na jednu stranu (malé hodnoty pro jednostrannou alternativu

$\mu < \mu_0$ nebo velké hodnoty pro alternativu $\mu > \mu_0$) budou svědčit spíše o platnosti alternativní hypotézy. Kritické obory tedy musejí mít tvar závislý na alternativní hypotéze a jsou shrnuty v tabulce 3.4. Ještě určíme p-hodnoty pro jednotlivé alternativní hypotézy. Je-li t vypočtená hodnota testového kritéria a T náhodná veličina s rozdělením $t(n-1)$, dostáváme z definice p-hodnoty postupně (F je distribuční funkce Studentova rozdělení t)

- $\mu < \mu_0$: p-hodnota $= P(T \leq t) = F(t)$,
- $\mu > \mu_0$: p-hodnota $= P(T \geq t) = 1 - F(t)$,
- $\mu \neq \mu_0$: p-hodnota $= 2P(T \geq |t|) = 2\cdot[1 - F(|t|)]$.

Tab. 3.4 Test o střední hodnotě normálního rozdělení

H_0	H_1	Testové kritérium	Kritický obor
$\mu = \mu_0$ rozptyl známý	$\mu < \mu_0$	$U = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}$ $U \sim N(0,1)$	$W_\alpha = \{u; u \leq -u_{1-\alpha}\}$
	$\mu > \mu_0$		$W_\alpha = \{u; u \geq u_{1-\alpha}\}$
	$\mu \neq \mu_0$		$W_\alpha = \{u; \lvert u\rvert \geq u_{1-\alpha/2}\}$
$\mu = \mu_0$ rozptyl neznámý	$\mu < \mu_0$	$T = \frac{\bar{X} - \mu_0}{S}\sqrt{n}$ $T \sim t(n-1)$	$W_\alpha = \{t; t \leq -t_{1-\alpha}\}$
	$\mu > \mu_0$		$W_\alpha = \{t; t \geq t_{1-\alpha}\}$
	$\mu \neq \mu_0$		$W_\alpha = \{t; \lvert t\rvert \geq t_{1-\alpha/2}\}$

Příklad 3.8

Předpokládejme, že v příkladu 3.4 má automat zhotovit kilogramové balíčky cukru. Z 20 náhodně vybraných balíčků byl zjištěn průměrný obsah 0,977 kg a výběrová směrodatná odchylka 0,015 kg. Položíme si otázku, zda zjištěná průměrná hmotnost již znamená, že stroj balí (ve střední hodnotě) méně než jeden kilogram. Budeme tedy testovat hypotézu

$H_0: \mu = 1$ proti alternativě $H_1: \mu < 1$.

Pokud by nás zajímala otázka, zda je střední hodnota rovna 1 kg nebo není, zvolili bychom alternativní hypotézu

$H_1: \mu \neq 1$.

Připomeňme ještě, že vzhledem k pozorovanému průměrnému obsahu, který je menší než jeden kilogram, nemá v tomto případě smysl testovat alternativu

$H_1: \mu > 1$.

V takovém případě bychom nikdy nulovou hypotézu nezamítli.

Vzhledem k tomu, že hodnotu rozptylu hmotností balíčků neznáme, použijeme testové kritérium

$$t = \frac{\bar{x} - \mu_0}{s}\sqrt{n} = \frac{0{,}977 - 1}{0{,}015}\sqrt{20} = -6{,}857$$

a kritický obor má podle tabulky 3.4 tvar ($t_{0,95}(19) = 1,729$)

$W_{0,05} = \{t; t \le -t_{0,95}\} = (-\infty; -1,729\rangle$.

Hodnota testového kritéria je prvkem kritického oboru, proto na hladině významnosti 5 % zamítáme nulovou hypotézu ve prospěch alternativy, že střední obsah balíčku je menší než předepsaný jeden kilogram. Můžeme také říci, že pozorovaná průměrná hmotnost je statisticky významně menší než jeden kilogram, nebo že náhodný výběr odporuje nulové hypotéze o průměrné hmotnosti 1 kg. Stejné rozhodnutí dostaneme také pomocí p-hodnoty, neboť

p-hodnota $= P(T \le -6,857) = F(-6,857) = 7,6 \cdot 10^{-7}$,

kde náhodná veličina T má Studentovo rozdělení s 19 stupni volnosti a F je distribuční funkce tohoto rozdělení.

■

Někdy se setkáváme také s úlohou testovat hypotézu o variabilitě náhodné veličiny s normálním rozdělením. Test hypotézy o rozptylu σ^2 je opět založen na bodovém odhadu rozptylu S^2, kterým jsme se již zabývali. Platí-li nulová hypotéza

$$H_0: \sigma^2 = \sigma_0^2, \tag{3.50}$$

má testová statistika

$$G = \frac{(n-1)S^2}{\sigma_0^2} \tag{3.51}$$

rozdělení χ^2 s $n-1$ stupni volnosti. Test nulové hypotézy (3.50) proti jedné z alternativních hypotéz

$$H_1: \sigma^2 \ne \sigma_0^2, \quad H_1: \sigma^2 < \sigma_0^2, \quad H_1: \sigma^2 > \sigma_0^2$$

tedy můžeme založit na testovém kritériu (3.51). V případě oboustranné alternativy budou proti nulové hypotéze svědčit příliš malé nebo příliš velké hodnoty testového kritéria. Proto je v takovém případě třeba nalézt dva kvantily rozdělení a kritický obor konstruovat ve dvou částech tak, jak je uvedeno v tabulce 3.5. V případě alternativní hypotézy, že skutečný rozptyl je větší než hypotetický rozptyl, budou proti nulové hypotéze hovořit příliš velké hodnoty testového kritéria, neboť bodový odhad bude větší než by měl být za platnosti nulové hypotézy. Pro alternativní hypotézu, že skutečný rozptyl je menší než hypotetický, budou proti platnosti nulové hypotézy svědčit malé hodnoty testového kritéria.

Při výpočtu p-hodnoty nemůžeme využít symetrii rozdělení tak, jak jsme to udělali v předchozích testech. Je-li g vypočtená hodnota testového kritéria a X náhodná veličina s rozdělením $\chi^2(n-1)$, lze hledanou p-hodnotu určit pro jednostranné alternativní hypotézy ve tvaru (F je distribuční funkce rozdělení χ^2 s $n-1$ stupněm volnosti)

- $\sigma^2 < \sigma_0^2$: p-hodnota $= P(X \le g) = F(g)$,
- $\sigma^2 > \sigma_0^2$: p-hodnota $= P(X \ge g) = 1 - F(g)$.

Je-li alternativní hypotéza oboustranná, hledanou p-hodnotu najdeme jako dvojnásobek menší z pravděpodobností $P(X \le g)$ a $P(X \ge g)$, tedy

- $\sigma^2 \ne \sigma_0^2$: p-hodnota $= 2 \cdot \min[F(g), 1 - F(g)]$

Tab. 3.5 Test o rozptylu v normálního rozdělení

H_0	H_1	Testové kritérium	Kritický obor
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$G = \frac{(n-1)S^2}{\sigma^2}$	$W_\alpha = \{g; g \le \chi^2_\alpha\}$
	$\sigma^2 > \sigma_0^2$		$W_\alpha = \{g; g \ge \chi^2_{1-\alpha}\}$
	$\sigma^2 \ne \sigma_0^2$	$G \sim \chi^2(n-1)$	$W_\alpha = \{g; g \le \chi^2_{\alpha/2} \vee g \ge \chi^2_{1-\alpha/2}\}$

Příklad 3.9

Uvažujme problém z příkladu 3.4 o hmotnosti balíčků. Z 20 náhodně vybraných balíčků byla určena výběrová směrodatná odchylka 0,015. Pro balicí automat by směrodatná odchylka σ měla být maximálně 0,01 kg. Posoudíme, zda automat splňuje tento předpoklad.

Řešení

Chceme testovat nulovou hypotézu (přesnost balicího automatu je dostatečná)

$H_0: \sigma = 0,01$

proti alternativě, že variabilita je příliš vysoká, tj.

$H_1: \sigma > 0,01$.

Pokud chceme použít test o rozptylu v normálním rozdělení σ^2, převedeme tyto hypotézy na tvary

$H_0: \sigma^2 = 0,01^2 = 0,000\,1$,

$H_1: \sigma^2 > 0,000\,1$.

Použijeme testové kritérium (3.51) a vypočteme

$$g = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19 \cdot 0,015^2}{0,000\,1} = 42,75.$$

Z tabulky 3.5 dostáváme kritický obor ($\chi^2_{0,95}(19) = 30,143$)

$W_{0,05} = \{g; g \ge \chi^2_{0,95}\} = \langle 30,143; \infty)$.

Hodnota testového kritéria leží v kritickém oboru, $g = 42,75 \in W_{0,05}$, a tedy data odporují nulové hypotéze, že balicí automat splňuje požadavek na směrodatnou odchylku hmotnosti obsahu. Na hladině významnosti 0,05 tedy zamítáme nulovou hypotézu, že balicí zařízení splňuje požadavek přesnosti ve prospěch alternativy, že variabilita je příliš vysoká.

Ještě určíme p-hodnotu. Má-li náhodná veličina X rozdělení $\chi^2(19)$, pak lze hledanou hodnotu určit jako

p-hodnota $= P(X \geq 42{,}75) = 1 - F(42{,}75) = 0{,}001\,4.$

Pokud bychom chtěli významnost charakterizovat pouze hvězdičkami, uvedli bychom dvě hvězdičky. ■

Test o střední hodnotě na základě velkého výběru

Velmi často není možné předpokládat, že náhodný výběr pochází z normálního rozdělení. Pokud předpokládáme jiné pravděpodobnostní rozdělení, například Poissonovo nebo exponenciální, můžeme použít testy konstruované pro tato rozdělení. Další možností je použít nějaký obecný test nepředpokládající konkrétní rozdělení, takové testy se nazývají neparametrické (Weiss, 2012, Anděl, 1978). Pokud náhodný výběr je dostatečně velký na to, aby bylo možné použít centrální limitní větu, používáme **test o střední hodnotě ve velkém náhodném výběru.**

Zabýváme se platností nulové hypotézy

$$H_0: E(X) = E(X)_0,$$

proti některé z alternativ:

$$H_1: E(X) < E(X)_0, \quad H_1: E(X) > E(X)_0, \quad H_1: E(X) \neq E(X)_0. \tag{3.52}$$

Jako testové kritérium zvolíme statistiku

$$U = \frac{\sqrt{n}\left(\bar{X} - E(X)_0\right)}{S}, \tag{3.53}$$

která má za platnosti nulové hypotézy přibližně (asymptoticky) normované normální rozdělení. Průběh testu je shrnut v tabulce 3.6.

Tab. 3.6 Test o střední hodnotě ve velkém náhodném výběru

H_0	H_1	Testové kritérium	Kritický obor
$E(X) = E(X)_0$	$E(X) < E(X)_0$	$U = \frac{\sqrt{n}\left(\bar{X} - E(X)_0\right)}{S}$	$W_\alpha = \{u;\ u \leq -u_{1-\alpha}\}$
	$E(X) > E(X)_0$		$W_\alpha = \{u;\ u \geq u_{1-\alpha}\}$
	$E(X) \neq E(X)_0$	$U \approx N(0, 1)$	$W_\alpha = \{u;\ \lvert u\rvert \geq u_{1-\alpha/2}\}$

Příklad 3.10

Předpokládejme, že nás zajímají mzdy v určitém odvětví v České republice. U 50 náhodně vybraných osob byla určena průměrná mzda 28 436 Kč a dále průměr druhých mocnin mezd 856 327 302. Lze na základě tohoto náhodného výběru usoudit, že střední mzda v odvětví je aspoň 28 000 Kč?

Řešení

Náhodný výběr budeme považovat za velký, proto nevadí, že mzdy nemají normální rozdělení. Pro jejich modelování se používá například logaritmicko-normální rozdělení. Zvolíme hladinu významnosti 5 %.

Nejprve určíme bodový odhad S^2 rozptylu σ^2. Použijeme-li vztah (3.4), lze psát

$$s^2 = \frac{50}{50-1}\left(856\,327\,302 - 28\,436^2\right) = 48\,695\,108,$$

$$s = \sqrt{48\,695\,108} = 6978 \text{ (Kč)}.$$

Budeme testovat nulovou hypotézu

$H_0: E(X) = 28\,000,$

proti alternativě

$H_1: E(X) > 28\,000.$

Použijeme-li testové kritérium (3.53), dostaneme

$$u = \frac{\sqrt{50}\left(28\,436 - 28\,000\right)}{6978} = 0{,}442.$$

Podle tabulky 3.6 má kritický obor tvar

$W_{0,05} = \{u;\ u \geq u_{0,95}\} = \langle 1{,}645;\ \infty)$.

Hodnota testového kritéria není prvkem kritického oboru, proto nemůžeme zamítnout nulovou hypotézu, že střední mzda je (maximálně) rovna 28 000 Kč.

■

3.4.3 Testy o parametrech na základě dvou výběrů

Předpokládejme nyní, že naším cílem je porovnat úroveň zkoumané náhodné veličiny ve dvou populacích. Můžeme zkoumat, zda ochota zúčastnit se voleb je stejná pro muže a ženy (test o rovnosti parametrů pro dvě alternativní rozdělení), nebo zda muži a ženy mají stejnou úroveň mezd (test o rovnosti dvou středních hodnot).

Testy rovnosti středních hodnot ve dvou normálních rozděleních

Nejprve uvažujme dva náhodné výběry z normálního rozdělení, první náhodný výběr o rozsahu n_1 z rozdělení $N(\mu_1, \sigma_1^2)$, druhý o rozsahu n_2 z rozdělení $N(\mu_2, \sigma_2^2)$. Nejprve budeme testovat hypotézu o rovnosti středních hodnot, tedy nulovou hypotézu

$$H_0: \mu_1 = \mu_2, \tag{3.54}$$

proti jedné z alternativ

$$H_1: \mu_1 \neq \mu_2, \quad H_1: \mu_1 < \mu_2, \quad H_1: \mu_1 > \mu_2. \tag{3.55}$$

Střední hodnoty obyčejně porovnáváme rozdílem, proto nulovou hypotézu a alternativní hypotézy přepíšeme do tvaru

$$H_0: \mu_1 - \mu_2 = 0 \tag{3.56}$$

a

$$H_1: \mu_1 - \mu_2 \neq 0, \quad H_1: \mu_1 - \mu_2 < 0, \quad H_1: \mu_1 - \mu_2 > 0. \tag{3.57}$$

Z hlediska vzniku náhodných výběrů můžeme rozlišovat dvě základní možnosti – nezávislé a závislé náhodné výběry. **Nezávislé náhodné výběry** jsou tvořeny například mzdami mužů a žen, nebo cenou bytů v panelových a cihlových domech. **Závislé (párové) náhodné výběry** jsou prováděny na stejných statistických jednotkách, tedy je například zjišťována výkonost firmy před a po zavedení nové technologie nebo spotřeba auta před a po úpravě motoru směřující ke snížení spotřeby.

Pokud chceme zvolit test takový, aby jeho závěry byly co nejsilnější, je vhodné si položit tyto otázky:

- Jsou zkoumané náhodné výběry nezávislé nebo závislé (párové)?
- Známe rozptyly v obou základních souborech?
- Můžeme předpokládat, že rozptyly jsou stejné?

Jestliže se pokusíme odpovědět na tyto otázky, umožní nám to volit optimální test s největší silou. Využití informace, kterou o analyzovaných datech máme, je důležité zvláště v případě malých náhodných výběrů, kdy máme o neznámých charakteristikách poměrně málo informace pocházející z dat. Pokud na otázky neodpovíme, je vždy možné použít nezávislé výběry s neznámými a různými rozptyly. Musíme tak ale dělat s vědomím, že tyto výsledky nemusí být v našem případě optimální, i když obyčejně (aspoň pro relativně velké výběry) vycházejí obdobně jako optimální volba.

Předpokládejme nejdříve, že náhodné výběry jsou nezávislé. Označíme aritmetický průměr $\bar{X}_1$ a směrodatnou odchylku S_1 v prvním náhodném výběru a $\bar{X}_2$ a S_2 ve druhém výběru. Pokud oba rozptyly σ_1^2 a σ_2^2 známe, pro test hypotézy (3.54) (respektive (3.56)) můžeme využít statistiku

$$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \tag{3.58}$$

která má za platnosti nulové hypotézy normované normální rozdělení.

Pokud rozptyly nejsou známé, budeme test nazývat **dvouvýběrovým *t*-testem**, neboť testová statistika bude mít Studentovo rozdělení t. Pokud můžeme předpokládat, že neznámé rozptyly v obou rozděleních jsou stejné, a tedy platí $\sigma_1^2 = \sigma_2^2$, společný rozptyl v obou náhodných výběrech je možné odhadnout statistikou

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

a test založit na testové statistice

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{3.59}$$

která má Studentovo rozdělení t s $n_1 + n_2 - 2$ stupni volnosti.

Poslední a nejobecnější možností jsou neznámé a různé rozptyly. V tomto případě lze použít testové kritérium

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \tag{3.60}$$

které má za platnosti nulové hypotézy přibližně Studentovo rozdělení. Stupně volnosti nelze vyjádřit přesně, lze je odhadnout například veličinou (Anděl, 2003)

$$\nu = \frac{\left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2}{n_2}\right)^2}. \tag{3.61}$$

Hodnota zlomku je náhodná veličina, neboť závisí na výběrových rozptylech, navíc není přirozené číslo. Můžeme ale říci, že její hodnota je vždy menší než $n_1 + n_2 - 2$, což je počet stupňů volnosti pro test s předpokladem rovných rozptylů. Výše zmíněné testy jsou shrnuty v tabulce 3.7.

Tab. 3.7 Test o rovnosti středních hodnot v normálním rozdělení

H_0	H_1	Testové kritérium	Kritický obor
$\mu_1 = \mu_2$ známé rozptyly	$\mu_1 < \mu_2$ $\mu_1 > \mu_2$ $\mu_1 \neq \mu_2$	$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $U \sim N(0, 1)$	$W_\alpha = \{u; u \leq -u_{1-\alpha}\}$ $W_\alpha = \{u; u \geq u_{1-\alpha}\}$ $W_\alpha = \{u; \lvert u\rvert \geq u_{1-\alpha/2}\}$
$\mu_1 = \mu_2$ neznámé stejné rozptyly	$\mu_1 < \mu_2$ $\mu_1 > \mu_2$ $\mu_1 \neq \mu_2$	$T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $T \sim t(n_1 + n_2 - 2)$	$W_\alpha = \{t; t \leq -t_{1-\alpha}\}$ $W_\alpha = \{t; t \geq t_{1-\alpha}\}$ $W_\alpha = \{t; \lvert t\rvert \geq t_{1-\alpha/2}\}$
$\mu_1 = \mu_2$ neznámé různé rozptyly	$\mu_1 < \mu_2$ $\mu_1 > \mu_2$ $\mu_1 \neq \mu_2$	$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ $T \approx t(\nu)$	$W_\alpha = \{t; t \leq -t_{1-\alpha}\}$ $W_\alpha = \{t; t \geq t_{1-\alpha}\}$ $W_\alpha = \{t; \lvert t\rvert \geq t_{1-\alpha/2}\}$

Příklad 3.11

Ve dvou firmách bylo vybráno 12 a 15 zaměstnanců a byl zjištěn průměrný věk. Byly určeny průměrné věky (v letech) 42,3 a 37,2 a výběrové směrodatné odchylky 8,2 a 5,5. Je možné na základě dat usoudit, že střední věk pracovníků první firmy je vyšší než ve druhé firmě?

Řešení

Zvolíme X_1 věk v první firmě a X_2 věk ve druhé firmě. Potom máme testovat hypotézu

H_0: $\mu_1 = \mu_2$ proti alternativě H_1: $\mu_1 > \mu_2$.

Víme, že $n_1 = 12$, $\bar{x}_1 = 42{,}3$, $s_1 = 8{,}2$ a $n_2 = 15$, $\bar{x}_2 = 37{,}2$, $s_2 = 5{,}5$. Vzhledem k tomu, že se jedná o dvě různé firmy, jsou výběry nezávislé, proto použijeme nějaký dvouvýběrový test. Vzhledem k podobným pozorovaným směrodatným odchylkám budeme předpokládat, že oba rozptyly jsou stejné. Určíme tedy společný odhad rozptylu jako

$$s^2 = \frac{(12-1)\cdot 8{,}2^2 + (15-1)\cdot 5{,}5^2}{12+15-2} = 46{,}53,$$

a z toho $s = \sqrt{46{,}53} = 6{,}82$.

Potom podle (3.59) je (zvolíme $\alpha = 0{,}05$)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{42{,}3-37{,}2}{6{,}82\sqrt{\frac{1}{12}+\frac{1}{15}}} = 1{,}93.$$

Použijeme-li kritický obor z tabulky 3.7, dostaneme (stupně volnosti Studentova rozdělení jsou 12 + 15 – 2 = 25)

$$W_{0,05} = \{t;\, t \geq t_{0,95}\} = \langle t_{0,95}, \infty) = \langle 1{,}708; \infty),$$

a hodnota testového kritéria 1,93 je prvkem kritického oboru. Proto zamítáme hypotézu o rovnosti středních hodnot věku ve prospěch alternativy, že v první firmě je střední hodnota větší.

■

Uvažujme nyní náhodný výběr o rozsahu n z dvourozměrného normálního rozdělení (X_1, X_2), kde X_1 má rozdělení $N(\mu_1, \sigma_1^2)$, a X_2 má rozdělení $N(\mu_2, \sigma_2^2)$. Představme si, že porovnáváme výkonnost pracovníků před a po určitém školení a máme pro vybrané pracovníky informaci o jejich výkonosti před i po absolvování školení. Můžeme tedy pro každého pracovníka vyčíslit změnu, ke které došlo po absolvování školení. V takovém případě se test rovnosti středních hodnot μ_1 a μ_2 v obou výběrech nazývá **párový *t*-test**. Problém konstrukce testu hypotézy (3.56) proti některé z alternativ (3.57) v případě párových náhodných výběrů převedeme na test o střední hodnotě v jednom výběru tak, že použijeme náhodný výběr zmíněných rozdílů $D_i = X_{1i} - X_{2i}$,

$i = 1, 2, \ldots, n$. Veličiny $D_1, D_2, \ldots, D_n$ tvoří náhodný výběr z normálního rozdělení náhodné veličiny $D = X_1 - X_2$ a platí $E(D) = \mu_D = \mu_1 - \mu_2$. Test nulové hypotézy lze tedy převést na test hypotézy o střední hodnotě rozdílů H_0: $\mu_D = 0$ proti vybrané alternativě. Použijeme (obdobně jako v (3.49)) testovou statistiku

$$T = \frac{\sqrt{n}\bar{D}}{S_D}, \tag{3.62}$$

kde $\bar{D}$ je průměrný rozdíl (a platí $\bar{D} = \bar{X}_1 - \bar{X}_2$) a S_D je výběrová směrodatná chyba rozdílu. Testové kritérium (3.62) má za platnosti nulové hypotézy Studentovo rozdělení s $n-1$ stupni volnosti. Z testového kritéria je patrné, že v případě oboustranné alternativy hodnoty testového kritéria blízké nule hovoří ve prospěch nulové hypotézy, hodnoty vzdálenější v její neprospěch. Provedení testu je v tabulce 3.8.

Tab. 3.8 Párový *t*-test

H_0	H_1	Testové kritérium	Kritický obor
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$T = \frac{\sqrt{n}\bar{D}}{S_D}$ $T \sim t(n-1)$	$W_\alpha = \{t;\, t \leq -t_{1-\alpha}\}$
	$\mu_1 > \mu_2$		$W_\alpha = \{t;\, t \geq t_{1-\alpha}\}$
	$\mu_1 \neq \mu_2$		$W_\alpha = \{t;\, \lvert t\rvert \geq t_{1-\alpha/2}\}$

Test o rovnosti rozptylů ve dvou výběrech z normálního rozdělení

Učinit předpoklad o stejných rozptylech ve dvou normálních rozděleních můžeme na základě úvahy, porovnáním velikosti výběrových rozptylů nebo můžeme shodu rozptylů testovat. ***F*-test** umožňuje porovnat velikost rozptylů ve dvou nezávislých výběrech z normálních rozdělení, používáme ho tedy kdykoliv, kdy nechceme porovnávat úroveň náhodných výběrů, ale zajímá nás pouze porovnání variability. V další části budeme testovat rozptyly náhodné veličiny ve více než dvou náhodných výběrech.

Na rozdíl od středních hodnot, které porovnáváme rozdílem, rozptyly porovnáváme podílem. Proto nulovou hypotézu o shodě rozptylů

$$H_0: \sigma_1^2 = \sigma_2^2 \tag{3.63}$$

přepíšeme na tvar

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1.$$

Tři alternativy, které budeme uvažovat, pak jsou

$$H_1: \sigma_1^2 \neq \sigma_2^2, \quad H_1: \sigma_1^2 < \sigma_2^2, \quad H_1: \sigma_1^2 > \sigma_2^2. \tag{3.64}$$

Test můžeme založit na testovém kritériu

$$F = \frac{S_1^2}{S_2^2}, \tag{3.65}$$

které má za platnosti nulové hypotézy Fisherovo-Snedecorovo F rozdělení se stupni volnosti $n_1 - 1$ a $n_2 - 1$. Pokud platí nulová hypotéza, hodnota testového kritéria by se měla pohybovat v okolí jedničky, hodnota podílu výběrových rozptylů (3.65) blízká nule nebo naopak výrazně větší než jedna bude svědčit proti platnosti nulové hypotézy. Postup testu je shrnut v tabulce 3.9.

Tab. 3.9 Test o rovnosti rozptylů ve dvou výběrech z normálního rozdělení

H_0	H_1	Testové kritérium	Kritický obor
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$ $F \sim F(n_1 - 1, n_2 - 1)$	$W_\alpha = \{F; F \leq F_{1-\alpha}\}$ $W_\alpha = \{F; F \geq F_{1-\alpha}\}$ $W_\alpha = \{F; F \leq F_{\alpha/2} \vee F \geq F_{1-\alpha/2}\}$

Příklad 3.12

Předpokládejme, že pro dvě doby (v minutách) výroby určitého výrobku ve dvou dílnách byly nalezeny výběrové rozptyly 2,4 a 3,5 min^2 určené pro 25 náhodně vybraných výrobků v první dílně a 30 náhodně vybraných výrobků ve druhé dílně. Je možné předpokládat, že variabilita doby výroby v obou dílnách je stejná?

Řešení

Máme $n_1 = 25$, $n_2 = 30$, $S_1^2 = 2{,}4$ a $S_2^2 = 3{,}5$. Testové kritérium F má podle (3.65) tvar

$$F = \frac{2{,}4}{3{,}5} = 0{,}686.$$

Zvolíme-li $\alpha = 0{,}1$, je $\alpha/2 = 0{,}05$ a $1 - \alpha/2 = 0{,}95$. Použijeme tabulku 3.9 a v tabulkách kvantilů nebo pomocí softwaru najdeme, že 5% kvantilem Fisherova-Snedecorova rozdělení s 24 ($n_1 - 1 = 25 - 1$) a 29 ($n_2 - 1 = 30 - 1$) stupni volnosti je 0,514 a 95% kvantilem je hodnota 1,901. Kritický obor $W_{0,1}$ je potom složen ze dvou intervalů $(0; 0{,}514\rangle$ a $\langle 1{,}901; \infty)$, vypočtená hodnota testového kritéria 0,686 je prvkem oboru přijetí (0,514; 1,901). Na hladině významnosti 0,1 tudíž nulovou hypotézu nezamítáme. Můžeme tedy předpokládat, že variabilita mezi dobami nutnými k výrobě výrobku, popsaná rozptylem, je v obou dílnách stejná. Všimněme si, že test neříká nic o průměrné nebo střední době výroby, která může být v obou dílnách odlišná. Zabývá se pouze kolísáním kolem střední hodnoty bez ohledu na její hodnotu. ■

Test o rovnosti středních hodnot ve dvou nezávislých výběrech

Předpokládejme dále, že stejně jako v předchozím výkladu máme k dispozici dva velké nezávislé náhodné výběry, výběr o rozsahu n_1 z rozdělení náhodné veličiny X_1 a výběr o rozsahu n_2 z rozdělení náhodné veličiny X_2. Nepředpokládáme ovšem, že

tyto veličiny mají normální rozdělení. Cílem je posoudit, zda oba výběry pocházejí z rozdělení se stejnou střední hodnotou, tedy testovat hypotézu o rovnosti středních hodnot $E(X_1)$ a $E(X_2)$. Označíme opět aritmetický průměr a výběrový rozptyl v prvním výběru $\bar{X}_1$ a S_1^2 a ve druhém výběru $\bar{X}_2$ a S_2^2.

Stejně jako v (3.54) budeme testovat hypotézu

$$H_0: E(X_1) = E(X_2)$$

proti některé z alternativních hypotéz:

$$H_1: E(X_1) \neq E(X_2), \quad H_1: E(X_1) < E(X_2), \quad H_1: E(X_1) > E(X_2).$$

Test lze založit na statistice (3.60), která má za platnosti nulové hypotézy asymptoticky (pro oba výběry dostatečně velké pro použití centrální limitní věty) normované normální rozdělení. Postup testu je shrnut v tabulce 3.10.

Tab. 3.10 Test rovnosti středních hodnot ve velkých výběrech

H_0	H_1	Testové kritérium	Kritický obor
$E(X_1) = E(X_2)$	$E(X_1) < E(X_2)$ $E(X_1) > E(X_2)$ $E(X_1) \neq E(X_2)$	$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ $U \approx N(0, 1)$	$W_\alpha = \{u; u \leq -u_{1-\alpha}\}$ $W_\alpha = \{u; u \geq u_{1-\alpha}\}$ $W_\alpha = \{u; \lvert u \rvert \geq u_{1-\alpha/2}\}$

Příklad 3.13

Předpokládejme, že cílem je porovnat doby do poruchy přístroje, který vyrábějí dva výrobci. Od prvního výrobce bylo vybráno 30 přístrojů a byla zjištěna průměrná doba do poruchy 20 měsíců a výběrová směrodatná odchylka 5 měsíců. U druhého výrobce bylo náhodně vybráno 50 výrobků a zjištěna průměrná doba do poruchy 22 měsíců a výběrová směrodatná odchylka 6 měsíců.

Řešení

Ze zadání dostaneme $n_1 = 30$, $\bar{x}_1 = 20$, $s_1 = 5$, $\bar{x}_2 = 22$, $n_2 = 50$, $s_2 = 6$ a zvolíme nulovou hypotézu a alternativu

$$H_0: E(X_1) = E(X_2),$$
$$H_1: E(X_1) < E(X_2).$$

Zajímá nás tedy, zda průměrná doba do poruchy přístrojů od prvního výrobce 20 měsíců je (statisticky významně) kratší než průměrná doba 22 měsíců, zjištěná u druhého výrobce.

Testové kritérium nabývá hodnoty

$$u=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}=\frac{20-22}{\sqrt{\frac{5^2}{30}+\frac{6^2}{50}}}=-1,288$$

a podle tabulky 3.10 dostaneme kritický obor

$$W_{0,05}=\{u;\ u\le -u_{1-0,05}\}=(-\infty;\ -1,645\rangle.$$

Vzhledem k tomu, že platí $u=-1,288\notin W_{0,05}$, nezamítáme na hladině významnosti $\alpha=0,05$ nulovou hypotézu. Rozdíl dva měsíce v průměrné době do poruchy není statisticky významný a nemůžeme usoudit, že střední doba do poruchy přístroje druhého výrobce je delší než od prvního výrobce.

■

Test o rovnosti parametrů ve dvou velkých nezávislých výběrech z alternativního rozdělení

Předpokládejme dále, že na základě dvou nezávislých velkých náhodných výběrů z alternativních rozdělení s parametry π_1 a π_2 chceme posoudit, zda tyto parametry jsou shodné. Například nás zajímá, zda procento vadných výrobků je stejné u dvou výrobců nebo zda ženy a muži volí stejně často určitou stranu. Stejně jako v předchozích kapitolách označíme n_1 rozsah prvního náhodného výběru, M_1 absolutní a P_1 relativní četnost v prvním náhodném výběru a n_2 rozsah náhodného výběru, M_2 a P_2 absolutní a relativní četnost ve druhém výběru.

Za platnosti nulové hypotézy, že parametry obou rozdělení jsou stejné, pocházejí oba náhodné výběry ze stejného alternativního rozdělení s parametrem $\pi=\pi_1=\pi_2$. Proto spojíme oba náhodné výběry do jednoho výběru o n_1+n_2 pozorováních. Společnou hodnotu neznámého parametru odhadneme relativní četností P ve spojeném výběru, tedy hodnotou

$$P=\frac{M_1+M_2}{n_1+n_2}=\frac{n_1P_1+n_2P_2}{n_1+n_2}=\frac{n_1}{n_1+n_2}P_1+\frac{n_2}{n_1+n_2}P_2.$$

Všimněme si, že P je váženým aritmetickým průměrem relativních četností v obou výběrech s vahami rovnými rozsahům výběrů. Bodový odhad P_1-P_2 rozdílu parametrů $\pi_1-\pi_2$ je konzistentní a nezkreslený. Pro test hypotézy

$$H_0:\ \pi_1=\pi_2$$

proti jedné z alternativních hypotéz

$$H_1:\ \pi_1\neq\pi_2,\quad H_1:\ \pi_1<\pi_2,\quad H_1:\ \pi_1>\pi_2 \tag{3.66}$$

použijeme statistiku

$$U=\frac{P_1-P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

která má za platnosti nulové hypotézy pro oba velké náhodné výběry asymptoticky normované normální rozdělení. Za platnosti nulové hypotézy se budou hodnoty kritéria nacházet v okolí nuly, odchylka na obě strany (v případě oboustranné alternativy) nebo na jednu stranu (v případě jednostranných alternativ) hovoří v neprospěch platnosti nulové hypotézy, že pravděpodobnosti jsou stejné.

Tab. 3.11 Test o rovnosti parametrů ve dvou výběrech z alternativních rozdělení

H_0	H_1	Testové kritérium	Kritický obor
$\pi_1=\pi_2$	$\pi_1<\pi_2$	$U=\dfrac{P_1-P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $U\approx N(0,1)$	$W_\alpha=\{u;\ u\le -u_{1-\alpha}\}$
	$\pi_1>\pi_2$		$W_\alpha=\{u;\ u\le -u_{1-\alpha}\}$
	$\pi_1\neq\pi_2$		$W_\alpha=\{u;\ \lvert u\rvert\ge u_{1-\alpha/2}\}$

Příklad 3.14

Vedení města zajímá, zda muži a ženy mají stejný názor na určité chystané opatření. Jestliže souhlasí 150 žen z 200 dotázaných a 210 mužů z 260 dotázaných, je možné usoudit, že muži i ženy mají stejný názor?

Řešení

Ze zadání víme, že $n_1=200$, $m_1=150$, $p_1=0,75$ a $n_2=260$, $m_2=210$, $p_2=0,81$.

Test nulové hypotézy

$$H_0:\ \pi_1=\pi_2$$

proti alternativní hypotéze

$$H_1:\ \pi_1\neq\pi_2$$

provedeme podle tabulky 3.11. Nejprve určíme odhad společné pravděpodobnosti, že občan souhlasí

$$n_1+n_2=460,\ m_1+m_2=360,\quad p=\frac{360}{460}=0,783.$$

Dále je třeba nalézt hodnotu testového kritéria

$$u=\frac{0,75-0,81}{\sqrt{0,783(1-0,783)\left(\frac{1}{200}+\frac{1}{260}\right)}}=-1,55.$$

Volíme-li $\alpha=0,05$, pro kritický obor platí

$$W_{0,05}=(-\infty;\ -1,96\rangle\cup\langle 1,96;\ \infty)$$

a pro vypočtenou hodnotu testového kritéria platí $u=-1,55\notin W_{0,05}$. Nemůžeme tedy zamítnout nulovou hypotézu, že muži a ženy mají stejný názor na chystané opatření. Také říkáme, že data neodporují nulové hypotéze. Pro p-hodnotu je

$$p\text{-hodnota}=2\cdot P(U\le -1,55)=2\cdot\Phi(-1,55)=0,121$$

a rozhodnutí je stejné, neboť platí p-hodnota $>0,05$.

■

3.4.4 Testy o parametrech ve více než dvou nezávislých výběrech

V této části budeme zkoumat případ, kdy máme více než dva náhodné výběry, budeme předpokládat, že jich je k ($k > 2$). Předpokládejme, že pozorování v j-tém náhodném výběru ($j = 1, 2, \ldots, k$) o rozsahu n_j budou mít rozdělení $N(\mu_j, \sigma_j^2)$ a označíme je $X_{j1}, X_{j2}, \ldots, X_{jn_j}$. Tato situace může nastat například tak, že při šetření příjmů rozdělíme respondenty podle vzdělání do čtyř skupin (základní, střední, úplné střední, vysokoškolské). V takovém případě bude $k = 4$ a můžeme se ptát, zda střední příjem bude stejný pro všechna vzdělání (test o shodě středních hodnot) nebo zda variabilita příjmů je stejná ve všech skupinách (test o shodě rozptylů).

Test o rovnosti středních hodnot

Zabývejme se nejprve testem hypotézy o rovnosti středních hodnot. Metoda pracuje s variabilitou proměnné X a s jejím rozkladem, proto ji nazýváme analýzou rozptylu. V tomto textu ukážeme pouze její jednofaktorovou verzi. Faktorem je proměnná, která přiřazuje statistické jednotky do jednotlivých náhodných výběrů. Používáme také název ANOVA, který pochází z angličtiny a je složen z počátečních písmen anglického názvu **AN**alysis **O**f **VA**riance.

Budeme předpokládat, že variabilita ve všech náhodných rozděleních je stejná, a tedy platí $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma^2$, kde jsme σ^2 označili společnou hodnotu rozptylu. Za tohoto předpokladu budeme testovat nulovou hypotézu, že střední hodnoty μ_j, $j = 1, 2, \ldots, k$ jsou stejné (všechny výběry pocházejí ze stejného normálního rozdělení)

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k, \tag{3.67}$$

proti alternativní hypotéze, že aspoň jedna střední hodnota je jiná. Bodovými odhady středních hodnot v jednotlivých normálních rozděleních jsou aritmetické průměry $\bar{X}_j$, $j = 1, 2, \ldots, k$. Za platnosti nulové hypotézy mají všechna normální rozdělení stejnou střední hodnotu, kterou můžeme odhadnout společným průměrem $\bar{X}$ určeným ze spojeného výběru o rozsahu $n_1 + n_2 + \ldots + n_k = n$.

Nejprve určíme celkový součet čtverců odchylek ve spojeném náhodném výběru o rozsahu n bez ohledu na různé náhodné výběry

$$S_x = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(X_{ji} - \bar{X}\right)^2.$$

Tento součet rozdělíme, stejně jako v 1. kapitole, na **součet čtverců mezi skupinami**

$$S_{x.m} = \sum_{j=1}^{k} n_j \left(\bar{X}_j - \bar{X}\right)^2,$$

a **součet čtverců uvnitř skupin**

$$S_{x.v} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(X_{ji} - \bar{X}_j\right)^2.$$

Obdobně jako v první kapitole v části věnované rozkladu rozptylu lze ukázat, že pro tyto tři součty platí vztah

$$S_x = S_{x.m} + S_{x.v}.$$

Pokud platí nulová hypotéza, aritmetické průměry v jednotlivých náhodných výběrech jsou podobné a blízké celkovému průměru, hodnota součtu čtverců mezi skupinami by měla být malá a součet čtverců uvnitř skupin srovnatelný s celkovým součtem čtverců. Test založíme na testové statistice

$$F = \frac{\frac{S_{x.m}}{k-1}}{\frac{S_{x.v}}{n-k}}, \tag{3.68}$$

která má za platnosti nulové hypotézy Fisherovo-Snedecorovo rozdělení s $k-1$ a $n-k$ stupni volnosti. Proti platnosti nulové hypotézy hovoří velké hodnoty testového kritéria, kritický obor má tedy tvar

$$W_\alpha = \{F; F \geq F_{1-\alpha}\} = \langle F_{1-\alpha}, \infty). \tag{3.69}$$

Průběh testu se obyčejně shrnuje v tabulce ANOVA, která je obsahem tabulky 3.12. Pro správné použití testu by měly být splněny oba výše zmíněné předpoklady (o normálním rozdělení a o stejných rozptylech v jednotlivých náhodných výběrech). Pro posouzení normality můžeme použít různé grafické metody, chí-kvadrát test dobré shody (část 3.4.5) nebo Kolmogorovův-Smirnovův test dobré shody (Malá, 2016, Anděl, 1978). Rovnost rozptylů lze testovat například Bartlettovým testem (viz dále). Nyní se spokojíme s přibližným kritériem a použijeme test ANOVA v případě, že podíl největší a nejmenší směrodatné odchylky je maximálně 3. V případě nesplnění předpokladů je možné použít vážení pozorování nebo test, který nepředpokládá normální rozdělení nebo stejné rozptyly (Hebák, 2013).

Tab. 3.12 Tabulka analýzy rozptylu

Zdroj variability	Stupně volnosti	Součet čtverců	Průměrný čtverec	F	p-hodnota
Mezi skupinami	$k-1$	$S_{x.m}$	$S_{x.m/(k-1)}$	podle (3.68)	
Uvnitř skupin	$n-k$	$S_{x.v}$	$S_{x.v/(n-k)}$		
Součet	$n-1$	S_x			

V případě, že zamítneme nulovou hypotézu o rovnosti středních hodnot, je přirozené položit si otázku, které střední hodnoty jsou stejné a které rozdílné. Existuje více možností, jak posuzovat rozdíly mezi střední hodnotami v k skupinách. V tomto textu ukážeme jen jeden postup, který posuzuje všechny možné dvojice skupin a používá upravený dvouvýběrový t-test z předchozí části. Metoda LSD (**L**east **S**ignificant **D**istance) pro výpočet testového kritéria využívá předpoklad o stejných rozptylech ve

všech skupinách a pro všechny dvojice skupin používá společnou směrodatnou odchylku s, určenou z analýzy rozptylu

$$s=\sqrt{\frac{S_{x.v}}{n-k}}.$$

Pro každou dvojici skupin j a l $(1 \le j \le l \le k)$ určíme směrodatnou odchylku rozdílu aritmetických průměrů podle vzorce

$$s\cdot\sqrt{\frac{1}{n_j}+\frac{1}{n_l}}.$$

Dále najdeme nejmenší významnou odchylku pro rozdíl průměrů těchto dvou skupin $|\bar{x}_j-\bar{x}_l|$ jako (počet stupňů volnosti pro kvantil Studentova rozdělení je $n-k$)

$$t_{1-\alpha/2}\, s\cdot\sqrt{\frac{1}{n_j}+\frac{1}{n_l}}. \tag{3.70}$$

Pokud je absolutní rozdíl obou průměrů větší než tato hodnota, usoudíme, že střední hodnoty μ_j a μ_l jsou rozdílné a $\mu_j-\mu_l\neq 0$. Všimněme si, že dvojic z k skupin je $\binom{k}{2}$ a toto číslo může být poměrně velké. Například pro $k=3$ jsou dvojice tři, pro $k=4$ jich je 6 a pro $k=5$ dostaneme 10 dvojic.

Příklad 3.15

Předpokládejme, že majitel obchodu se zajímá o úspěšnost své vlastní značky výrobku (skupina vlastní) ve srovnání se třemi dalšími (skupiny výrobce A–C). Proto zjišťoval ve vybraných dnech tržbu za výrobky těchto značek. Charakteristiky skupin získané z náhodného výběru jsou uvedeny v tabulce 3.13.

Tab. 3.13 Charakteristiky výsledků značek

Skupina	Značka	n_i	$\bar{x}_j$	s_i^2	s_i	$n_j(\bar{x}_j-\bar{x})^2$
I	vlastní	20	17,39	3,71	1,94	3,07
II	výrobce A	25	19,59	2,10	1,45	81,58
III	výrobce B	25	16,44	3,49	1,87	45,07
IV	výrobce C	25	17,63	2,55	1,60	0,57
Součet		95	17,79	4,21	2,05	130,29

Charakteristiky jednotlivých skupin, obsažené v tabulce, nazveme **skupinové četnosti, skupinové průměry, skupinové rozptyly a směrodatné odchylky.** V posledním sloupci jsou uvedeny hodnoty, které vstupují do součtu čtverců mezi skupinami. V posledním řádku jsou uvedeny stejné charakteristiky pro celý náhodný výběr (počet pozorování, průměr, výběrový rozptyl a výběrová směrodatná odchylka pro všechna

data bez ohledu na výrobce). V posledním sloupci tohoto řádku je uveden součet sloupce, tedy hodnota $S_{x.m}$ (je možné ji porovnat s hodnotou v tabulce ANOVA 3.14). Posoudíme, zda je možné předpokládat, že střední tržba je stejná pro výrobky všech zkoumaných značek.

V tomto případě je $n=95$ a $k=4$. Podíl největší a nejmenší skupinové směrodatné odchylky je roven 1,94/1,45 = 1,33, tato hodnota je menší než 3, a proto můžeme analýzu rozptylu použít podle přibližného pravidla, pracujícího s podílem největšího a nejmenšího výběrového rozptylu. Formálně budeme testovat shodu rozptylů Bartlettovým testem v příkladu 3.16. Normalitu pozorování nebudeme testovat, omezíme se jen na krabičkový graf, který ukazuje symetrická rozdělení. Průběh testu hypotézy

H_0: zkoumané značky jsou stejně úspěšné

proti alternativě

H_1: značky nejsou stejně úspěšné,

je shrnut v tabulce 3.14.

Tab. 3.14 Tabulka analýzy rozptylu

Zdroj variability	Stupně volnosti	Součet čtverců	Průměrný čtverec	F	p-hodnota
Mezi skupinami	3	130,2835	43,4279	14,8620	$5{,}93\cdot10^{-8}$
Uvnitř skupin	91	265,9088	2,9221		
Součet	94	396,1923			

Použijeme-li výsledky z tabulky ANOVA, zamítneme nulovou hypotézu o rovnosti středních hodnot ve prospěch alternativy, že aspoň jedna střední hodnota je odlišná. Můžeme použít p-hodnotu a porovnat ji s hladinou významnosti 0,05. V případě takto malé minimální hladiny významnosti se někdy ve statistickém softwaru uvádí jen výsledek p-hodnota < 0,000 0. Stejný závěr dostaneme pomocí kritického oboru, neboť platí $F_{0,95}(3,91)=2{,}705$, a tedy podle (3.69) je

$W_{0,05}=\{F;\, F\ge 2{,}705\}=\langle 2{,}705;\,\infty)$

a platí $F=14{,}86\in W_{0,05}$.

Nyní se pomocí metody LSD pokusíme odpovědět na otázku, které značky se liší, a najdeme dvojice, kde jsou střední tržby patrně různé. Výsledky jsou shrnuty v tabulce 3.15 (kvantily Studentova rozdělení s 91 stupni volnosti nahradíme kvantily normovaného normálního rozdělení). Nejmenší významné rozdíly, určené podle (3.70), jsou rovny pro počty pozorování 20 a 25 (skupina I a některá ze skupin II–IV)

$$u_{1-0{,}05/2}\sqrt{2{,}922}\sqrt{\frac{1}{20}+\frac{1}{25}}=1{,}019$$

a pro stejné počty 25 (pro dvojici skupin z II–IV)

$$u_{1-0{,}05/2}\sqrt{2{,}922}\sqrt{\frac{1}{25}+\frac{1}{25}}=0{,}960.$$

Tab. 3.15 Výsledky metody LSD

Skupiny	rozdíl průměrů	LSD	významnost
I-II	2,198	1,019	ano
I-III	0,951	1,019	ne
I-IV	0,241	1,019	ne
II-III	3,149	0,960	ano
II-IV	1,957	0,960	ano
III-IV	1,192	0,960	ne

Všimněme si, že skupina II, definovaná produkcí výrobce A, je odlišná od všech ostatních (podle průměrných tržeb uvedených v tabulce 3.13 patrně úspěšnější). Z ostatních dvojic metoda neukazuje na odlišnou střední tržbu. ■

Test o rovnosti rozptylů

Při aplikaci statistických metod se můžeme setkat s nutností testovat předpoklad, že variabilita ve více než dvou normálních rozděleních je stejná. Pokud nechceme použít při analýze rozptylu jen přibližné pravidlo pro posouzení shody rozptylů, můžeme tento předpoklad testovat. V takovém případě používáme test o shodě rozptylů. Budeme testovat hypotézu, že všechny rozptyly jsou stejné, tedy že platí

$$H_0: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2,$$

proti alternativě, že aspoň jeden rozptyl je jiný.

Celkový počet pozorování ve všech výběrech označíme n $\left(n=\sum_{j=1}^{k} n_j\right)$ a společný rozptyl S^2, určený za platnosti nulové hypotézy

$$S^2 = \frac{1}{n-k}\sum_{j=1}^{k}(n_j-1)\,S_j^2.$$

K testu nulové hypotézy můžeme použít Bartlettův test. Testové kritérium má tvar

$$G = \frac{(n-k)\ln S^2 - \sum_{j=1}^{k}(n_j-1)\ln S_j^2}{1+\frac{1}{3(k-1)}\left(\sum_{j=1}^{k}\frac{1}{n_j-1}-\frac{1}{n-k}\right)} \qquad (3.71)$$

a asymptoticky (pro velké n) má χ^2 rozdělení s $k-1$ stupni volnosti. Proti nulové hypotéze hovoří velké hodnoty testového kritéria. Proto můžeme kritický obor vyjádřit jako

$$W_\alpha = \left\{g;\, g \geq \chi^2_{1-\alpha}\right\} = \left\langle \chi^2_{1-\alpha}, \infty\right).$$

Nyní test použijeme k posouzení oprávněnosti předpokladu o shodě rozptylů při použití jednofaktorové analýzy rozptylu v příkladu 3.15.

Příklad 3.16

Budeme testovat hypotézu, že skupinové rozptyly v příkladu 3.15 jsou stejné. Nejprve určíme společný rozptyl

$$s^2 = \frac{1}{n-k}\sum_{j=1}^{k}(n_j-1)s_j^2 = \frac{1}{95-4}\left(19\cdot 3{,}71+24\cdot 2{,}10+24\cdot 3{,}49+24\cdot 2{,}55\right) = 2{,}922.$$

Ze vztahu (3.71) určíme hodnotu testového kritéria

$$g = \frac{(n-k)\ln s^2 - \sum_{j=1}^{k}(n_j-1)\ln s_j^2}{1+\frac{1}{3(k-1)}\left(\sum_{j=1}^{k}\frac{1}{n_j-1}-\frac{1}{n-k}\right)} =$$

$$= \frac{91\cdot\ln 2{,}922 - (19\cdot\ln 3{,}71 + 24\cdot\ln 2{,}10 + 24\cdot\ln 3{,}49 + 24\cdot\ln 2{,}55)}{1+\frac{1}{3(4-1)}\left(\frac{1}{19}+\frac{1}{24}+\frac{1}{24}+\frac{1}{24}-\frac{1}{91}\right)} = 2{,}331.$$

Za platnosti nulové hypotézy má testové kritérium rozdělení $\chi^2(3)$ a kritický obor je (platí $\chi^2_{0,95}(3)=7{,}81$)

$$W_\alpha = \left\{g;\, g \geq \chi^2_{0,95}\right\} = \left\langle 7{,}81; \infty\right).$$

Hodnota testového kritéria není prvkem kritického oboru, proto nulovou hypotézu nezamítáme. Skutečně lze tedy předpokládat, že všechny rozptyly jsou stejné. ■

3.4.5 Chí-kvadrát test dobré shody

Často je třeba posoudit, zda **pozorované** (též **empirické**) **rozdělení** náhodného výběru odpovídá nějakému **teoretickému** (také **hypotetickému**) rozdělení pravděpodobností. Jde například o odpověď na otázky, zda věkové složení náhodně vybraných respondentů odpovídá známému rozdělení v celé populaci, zda rozdělení prodaných výrobků podle jednotlivých výrobců odpovídá předpokladu, nebo třeba zda hrací kostka je pravidelná nebo také spravedlivá. Jedním z testů, které umožňují zabývat se podobnými otázkami, je **chí-kvadrát test dobré shody**, kterým se budeme zabývat v této části. V matematické statistice test také používáme pro posouzení, zda předpoklad o konkrétním pravděpodobnostním rozdělení je akceptovatelný jako model pro náhodný výběr. V obou případech teoretické rozdělení volíme na základě teoretického rozboru problému, našich znalostí a zkušeností. Dalším využitím je posouzení, zda náhodný výběr vyhovuje předpokladům statistických metod, které chceme použít, například předpokladu normálního rozdělení pozorování.

Test dobré shody nevybere nejlepší nebo správné rozdělení, umožňuje pouze odpovědět na otázku, zda předpokládané rozdělení je přijatelným modelem pro data nebo jím není. Parametry rozdělení mohou být známé i neznámé, ve druhém případě tyto parametry musíme nejprve odhadnout.

Předpokládejme nejprve, že existuje $k > 2$ neslučitelných výsledků nějakého náhodného pokusu a stanovme teoretické hodnoty pravděpodobností těchto výsledků $\boldsymbol{\pi}_0 = (\pi_{0,1}, \pi_{0,2}, ..., \pi_{0,k})'$, kde $\sum_{j=1}^{k} \pi_{0,j} = 1$. Pokud bychom se zabývali hrací kostkou a možné výsledky jednoho hodu jako počet ok 1 až 6, je $k = 6$ a pokud je kostka spravedlivá, je $\boldsymbol{\pi}_0 = (1/6; 1/6; 1/6; 1/6; 1/6; 1/6)'$. Pokud bychom si předepsali, že výběrové šetření zákazníků má obsahovat 30 % osob do 20 let, 30 % osob 20–30 let, 20 % osob 30–40 let a zbytek nad 40 let, protože takové je známé věkové složení cílové populace, je výsledkem pokusu zařazení vybrané osoby do jedné z věkových skupin. Dostaneme tedy $k = 4$ a $\boldsymbol{\pi}_0 = (0{,}3; 0{,}3; 0{,}2; 0{,}2)'$.

Předpokládejme, že máme náhodný výběr o rozsahu n, obsahující informaci o výsledku n takových pokusů – v našem případě hodů kostkou nebo věku vybraných respondentů. Pro posouzení platnosti nulové hypotézy

H_0: náhodná veličina má rozdělení popsané pravděpodobnostmi $\boldsymbol{\pi}_0$

proti alternativě

H_1: rozdělení náhodné veličiny je jiné,

porovnáme **výběrové** (také **empirické**) rozdělení s teoretickým rozdělením četností. Nejprve určíme **pozorované** (také **empirické**) **četnosti** jednotlivých tříd $n_1, n_2, \ldots, n_k$, platí $n_1 + n_2 + \ldots + n_k = n$, a porovnáme je s **očekávanými** (také **teoretickými**) **četnostmi** za platnosti rozdělení, které snadno získáme jako $n\pi_{0,1}, n\pi_{0,2}, \ldots, n\pi_{0,k}$. Pokud platí nulová hypotéza, očekávané a pozorované četnosti by měly být podobné, velké odchylky hovoří proti platnosti nulové hypotézy, neboť v takovém případě je rozdělení náhodného výběru patrně jiné, než je předpokládané rozdělení. Pro každou třídu ($j = 1, 2, \ldots, k$) porovnáme četnosti n_j a $n\pi_{0,j}$ pomocí relativní statistiky

$$\frac{(n_j - n\pi_{0,j})^2}{n\pi_{0,j}}.$$

Součet těchto statistik pro všechny třídy

$$G = \sum_{j=1}^{k} \frac{(n_j - n\pi_{0,j})^2}{n\pi_{0,j}} \tag{3.72}$$

použijeme jako testové kritérium. Za platnosti nulové hypotézy má testové kritérium rozdělení χ^2 s $k - 1$ stupni volnosti. Kritický obor bude mít tvar

$$W_\alpha = \{g; g \geq \chi^2_{1-\alpha}\} = \langle \chi^2_{1-\alpha}, \infty \rangle. \tag{3.73}$$

P-hodnotu pro tento test určíme jako pravděpodobnost, že náhodná veličina s rozdělením $\chi^2(k-1)$ nabude hodnoty větší nebo rovné testovému kritériu. Tato pravděpodobnost je rovna

$$1 - F(g), \tag{3.74}$$

kde F je distribuční funkce rozdělení χ^2 s $k - 1$ stupni volnosti a g je vypočtená hodnota testového kritéria.

Chí-kvadrát test dobré shody je asymptotickým testem, pro jeho správné použití klademe předpoklad na očekávané četnosti. Doporučuje se použít tento test tehdy, pokud přibližně $n\pi_j \geq 5$, $j = 1, 2, \ldots, k$. Tuto podmínku lze zeslabit na předpoklad, že každá z očekávaných četností je větší nebo rovna jedné a navíc aspoň 90 % z nich (někdy jen 80 %) je větších nebo rovných 5.

Tab. 3.16 Chí-kvadrát test dobré shody

H_0 a H_1	Testové kritérium	Kritický obor
H_0: $\pi_j = \pi_{0,j}$ $j = 1, 2, \ldots, k$ H_1: rozdělení je jiné	$G = \sum_{j=1}^{k} \frac{(n_j - n\pi_{0,j})^2}{n\pi_{0,j}}$ $G \approx \chi^2(k-1)$	$W_\alpha = \{g; g \geq \chi^2_{1-\alpha}\}$

Příklad 3.17

Prodejce předpokládá, že pokud bude nabízet knihu ve třech formách: jako paperback, pevnou vazbu a e-knihu, zájem čtenářů bude rozdělen v poměru 40 %, 15 % a 45 %. Po nákupu 160 knih byly zjištěny počty u jednotlivých forem 70, 12 a 78. Byl počáteční předpoklad o rozložení zájmu správný?

Řešení

Vzhledem k tomu, že máme tři formy knihy, je $k = 3$. Nulová hypotéza říká, že rozdělení zájmu o knihu je popsáno vektorem pravděpodobností

$\boldsymbol{\pi}_0 = (0{,}4; 0{,}15; 0{,}45)'$.

Platí-li nulová hypotéza, měl by být celkový počet zájemců 160 rozdělen na 64 (40 % ze 160), 24 (15 % ze 160) a 72 (45 % ze 160), jak je uvedeno ve sloupci očekávané četnosti v tabulce 3.17. V posledním sloupci tabulky jsou uvedeny příspěvky jednotlivých typů knihy do testového kritéria podle (3.72) nebo tabulky 3.16. Zvolíme hladinu významnosti $\alpha = 0{,}05$, kritickou hodnotou podle tabulky 3.16 je 95% kvantil rozdělení χ^2 se dvěma stupni volnosti a platí $\chi^2_{0,95}(2) = 5{,}99$. Hodnota testového kritéria 7,063 je větší než tato kritická hodnota, zamítáme proto nulovou hypotézu na zvolené hladině významnosti. p-hodnotu získáme výpočtem pravděpodobnosti $1 - F(7{,}063) = 0{,}029$, kde F je distribuční funkce rozdělení χ^2 se dvěma stupni volnosti.

Podíváme-li se do tabulky 3.17 na očekávané a pozorované četnosti vyjádřené v procentech, můžeme říci, že o pevnou vazbu byl jen poloviční zájem, než předpokládaný. Tuto skutečnost vyjadřuje také největší hodnota ve sloupci G, odrážející neshodu očekávaných a pozorovaných četností.

Tab. 3.17 Výpočet testového kritéria k příkladu 3.17

Typ	Očekávané četnosti (%)	$\pi_{0,j}$	Pozorované četnosti	Pozorované četnosti (%)	Teoretické četnosti	g
Paperback	40	0,40	70	43,75	64	0,562 5
Pevná vazba	15	0,15	12	7,50	24	6,000 0
e-kniha	45	0,45	78	48,75	72	0,500 0
Součet	100	1	160	100	160	7,062 5

■

Dále ukážeme, jak je možné tento test použít pro ověření hypotézy, že pravděpodobnostní rozdělení, popsané distribuční funkcí F, je možné použít pro zkoumanou náhodnou veličinu X. Můžeme testovat vhodnost pravděpodobnostního rozdělení se zadanými (tedy známými) parametry nebo ze stejného náhodného výběru tyto parametry nejprve odhadnout a potom testovat shodu empirického a zvoleného rozdělení s odhadnutými parametry. Předpokládejme tedy opět, že pracujeme s náhodným výběrem o rozsahu n.

Nejprve uvažujme případ, kdy známe parametry teoretického rozdělení, takové rozdělení nazýváme **plně specifikovaným**. Budeme testovat hypotézu

H_0: náhodný výběr pochází ze zvoleného rozdělení s distribuční funkcí F,

proti alternativě

H_1: rozdělení je jiné.

Pokud náhodná veličina nabývá jen izolovaných hodnot, sestrojíme tabulku rozdělení četností, kterou jsme používali v první kapitole této knihy. Pokud je rozdělení spojité, a tedy se hodnoty jednotlivých pozorování neopakují, sestrojíme tabulku intervalového rozdělení četností podle první kapitoly. K tomu je třeba rozdělit množinu všech pozorovaných hodnot náhodné veličiny X do k nepřekrývajících se intervalů, ve tvaru $(-\infty, x_1\rangle$, $(x_1, x_2\rangle$, …, (x_{k-1}, ∞). Oběma postupy získáme v těchto tabulkách **pozorované četnosti** v jednotlivých řádcích tabulky. Dále určíme pravděpodobnosti jednotlivých hodnot náhodné veličiny nebo jejich intervalů $\pi_{0,j}$, $j = 1, 2, \ldots, k$, použijeme známou distribuční funkci hypotetického rozdělení a určíme pravděpodobnosti jednotlivých tříd. Pokud se jedná o diskrétní rozdělení a v řádcích jsou jednotlivé hodnoty náhodné veličiny (jako v příkladu 3.18), jsou teoretické pravděpodobnosti rovny pravděpodobnostní funkci. V případě intervalových tříd najdeme

$$\pi_{0,1} = F(x_1),$$
$$\pi_{0,j} = F(x_{j+1}) - F(x_j), \quad j = 1, 2, \ldots, k-2,$$
$$\pi_{0,k} = 1 - F(x_{k-1}).$$

Dále již porovnáme teoretické a pozorované četnosti pomocí chí-kvadrát testu dobré shody. Jako testové kritérium použijeme statistiku (3.72) a kritický obor (3.73), případně tabulku 3.16.

V případě, že žádné (nebo aspoň některé) parametry hypotetického rozdělení nejsou známy, je třeba je nejprve odhadnout a potom použít postup uvedený výše. Pouze je třeba zmenšit počet stupňů volnosti χ^2 rozdělení testového kritéria $k-1$ o počet odhadnutých parametrů.

Pokud nějaké hypotetické četnosti jsou menší než 5, můžeme sloučit sousední třídy tak, aby výsledné teoretické četnosti splňovaly předpoklad. V takovém případě jednoduše sečteme pozorované četnosti všech sloučených tříd a obdobně i hypotetické četnosti. Dále je třeba přepočítat příspěvek nové třídy do testového kritéria a upravit, v tomto případě snížit stupně volnosti hypotetického rozdělení.

Příklad 3.18

Majitel firmy předpokládá, že za jednu směnu dochází v průměru ke dvěma poruchám. Jestliže ve vybraných dnech pozoroval počty poruch uvedené v tabulce 3.18, je možné předpokládat, že počet poruch za směnu má Poissonovo rozdělení s parametrem 2?

Řešení

V tabulce 3.18 jsou uvedeny pozorované četnosti a také teoretické pravděpodobnosti, určené z hypotetického rozdělení. Z pravděpodobnostní funkce Poissonova rozdělení s parametrem $\lambda = 2$

$$P(x) = \mathrm{e}^{-2}\frac{2^x}{x!}, \quad x = 0, 1, \ldots,$$

dostaneme například

$$P(0) = \mathrm{e}^{-2}\frac{2^0}{0!} = \mathrm{e}^{-2} = 0{,}135,$$

$$P(1) = \mathrm{e}^{-2}\frac{2^1}{1!} = 2\mathrm{e}^{-2} = 0{,}271.$$

Pro nejvyšší počet poruch za směnu (4) je třeba uvažovat (tato třída zahrnuje všechny možné hodnoty větší nebo rovné čtyřem)

$$P(X \geq 4) = 1 - (F(0) + F(1) + F(2) + F(3)) = 0{,}143.$$

Teoretické četnosti dostaneme násobením teoretických pravděpodobností rozsahem náhodného výběru 50. Pro první dva řádky dostaneme $0{,}135 \cdot 50 = 6{,}767$ a $0{,}271 \cdot 50 = 13{,}534$. V posledním sloupci jsou učeny hodnoty, které vstupují do testového kritéria. Pro první dva řádky například dostáváme

$$\frac{(3 - 6{,}767)^2}{6{,}767} = 2{,}097,$$

$$\frac{(6-13{,}534)^2}{13{,}534}=4{,}194.$$

Testové kritérium je rovno součtu hodnot v posledním sloupci, tedy 11,499.

Tab. 3.18 Výpočet testového kritéria pro příklad 3.18

Řádek	Hodnota	Pozorovaná četnost	$\pi_{0,j}$	Teoretická četnost	g
1	0	3	0,135	6,767	2,097
2	1	6	0,271	13,534	4,194
3	2	18	0,271	13,534	1,474
4	3	11	0,180	9,022	0,433
5	4	12	0,143	7,144	3,301
Součet		50	1	50	11,499

Tabulka má 5 řádků ($k=5$), proto má testové kritérium za platnosti nulové hypotézy χ^2 rozdělení s $5-1=4$ stupni volnosti. Všechny teoretické četnosti jsou aspoň 5, můžeme proto použít chí-kvadrát test dobré shody. Kritický obor je podle tabulky 3.17

$$W_{0,05}=\{g;\, g\geq\chi^2_{1-0,05}\}=\langle\chi^2_{0,95},\infty)=\langle 9{,}488;\infty)$$

a platí $g\in W_{0,05}$. Zamítáme tedy hypotézu o tom, že počet poruch má Poissonovo rozdělení s parametrem 2. Zřejmě je poruch v průměru více než 2, neboť průměrný počet poruch je (podle tabulky 3.18)

$$\bar{x}=\frac{6+36+33+48}{50}=2{,}5.$$

■

Příklad 3.19

Uvažujme náhodný výběr hmotností balíčků z příkladu 3.4, obsahující 80 pozorování. Volíme případ většího rozsahu náhodného výběru, neboť menší rozsah náhodného výběru 20 pozorování není postačující pro smysluplné použití chí-kvadrát testu dobré shody, který je testem asymptotickým. Budeme testovat hypotézu, že tento náhodný výběr pochází z normálního rozdělení, aniž bychom určili hodnoty parametrů, tj.

H_0: náhodný výběr pochází z normálního rozdělení.

Nejprve je třeba nalézt parametry rozdělení, které budeme používat pro výpočet teoretických četností v testu dobré shody. Podle části 3.3 můžeme použít aritmetický průměr a výběrový rozptyl a uvažovat rozdělení $N(\bar{X}, S^2)$. Již byly určeny hodnoty $\bar{x}=0{,}977$ kg a $s=0{,}015$ kg, tedy použijeme rozdělení $N(0{,}977;\ 0{,}015^2)$. Pro intervalové rozdělení četností zvolíme 9 tříd, neboť například podle odmocninového pravidla pro počet tříd je $\sqrt{80}\doteq 8{,}9$. Všimněme si, že v případě tohoto testu již není třeba zachovávat všechny třídy stejně široké tak, jak jsme dělali u intervalového rozdělení četností v první kapitole.

Tab. 3.19 Výpočet testového kritéria pro chí-kvadrát test dobré shody

Interval	Dolní mez intervalu	Horní mez intervalu	Pozorovaná četnost	$\pi_{0,j}$	Teoretická četnost	g
1		0,96	12	0,141	11,248	0,050
2	0,96	0,965	8	0,085	6,803	0,211
3	0,965	0,97	8	0,108	8,667	0,051
4	0,97	0,975	8	0,124	9,949	0,382
5	0,975	0,98	7	0,129	10,289	1,051
6	0,98	0,985	12	0,120	9,587	0,607
7	0,985	0,99	8	0,101	8,049	0,000
8	0,99	1	13	0,128	10,237	0,746
9	1		4	0,065	5,170	0,265
Součet			80	1	80	3,553

Nejprve pro jednotlivé intervaly v tabulce musíme nalézt teoretické pravděpodobnosti, že náhodná veličina X s rozdělením $N(0{,}977;\ 0{,}015^2)$ nabude hodnoty z daného intervalu. Pro první interval dostáváme (Φ je distribuční funkce normovaného normálního rozdělení)

$$P(X\leq 0{,}96)=\Phi\left(\frac{0{,}96-0{,}977}{0{,}015}\right)=0{,}129,$$

pro druhý interval (a obdobně i pro další intervaly s výjimkou posledního) je

$$P(0{,}96<X\leq 0{,}965)=\Phi\left(\frac{0{,}965-0{,}977}{0{,}015}\right)-\Phi\left(\frac{0{,}96-0{,}977}{0{,}015}\right)=0{,}083.$$

Pro poslední interval je

$$P(X>1)=1-\Phi\left(\frac{1-0{,}977}{0{,}015}\right)=0{,}063.$$

Teoretické četnosti určíme násobením těchto pravděpodobností rozsahem výběru, tedy hodnotou 80. V prvním řádku určíme míru shody mezi pozorovanými a teoretickými četnostmi

$$\frac{(12-10{,}283)^2}{10{,}283}=0{,}287$$

a obdobně budeme pokračovat i v dalších řádcích. Hodnota testového kritéria je rovna 3,553. Tabulka má 9 řádků, odhadli jsme 2 parametry, a tedy testové kritérium má za platnosti nulové hypotézy χ^2 rozdělení s $9-1-2=6$ stupni volnosti. Zvolíme-li hladinu významnosti 0,05, dostaneme kritický obor

$$W_{0,05}=\{g;\, g\geq\chi^2_{1-0,05}\}=\langle\chi^2_{0,95},\ \infty)=\langle 12{,}592;\ \infty).$$

Hodnota testového kritéria leží v oboru přijetí, proto nezamítáme nulovou hypotézu. Můžeme tedy předpokládat, že hmotnosti balíčků lze popsat normálním rozdělením.

■