VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
ÚSTAV SOUDNÍHO INŽENÝRSTVÍ

Doc. RNDr. Zdeněk Karpíšek, CSc.

ELEMENTÁRNÍ STATISTICKÉ METODY

UČEBNÍ TEXT

Brno 2018 – první vydání

OBSAH

PŘEDMLUVA

Tento učební text je koncipován jako podpora výuky předmětu Matematické základy analýzy rizika (RSMAT) na Ústavu soudního inženýrství Vysokého učení technického v Brně v magisterském studijním programu Rizikové inženýrství. Účelem textu je seznámení studentů se základními statistickými metodami a možnostmi jejich aplikací při modelování a vyhodnocování reálných jevů v oblasti rizik v ekonomice, finančnictví, výrobě apod. Jde pouze o elementární statistické metody, které by měly být běžnou součástí bakalářských programů vysokých škol ekonomického a technického zaměření nebo dokonce i součástí výuky matematiky na středních školách, ale bohužel tak tomu mnohdy není. Obsahově je text orientován na zpracování jednorozměrných a dvourozměrných statistických souborů, základní analýzu časových řad a analýzu indexů.

Text má pouze přehledový charakter a obsahuje pouze základní pojmy a postupy statistických metod neindukčního, tj. popisného typu. Další informace z této oblasti najde student nejen na www stránkách, ale také v literatuře uvedené v závěrečné části. Dodatek vlastního textu tvoří základní přehled elementů teorie pravděpodobnosti. Každá kapitola i dodatek jsou zakončeny kontrolními otázkami pro vlastní ověření získaných znalostí.

Děkuji všem, kteří mně pomohli svými připomínkami a radami pro přípravu tohoto učebního textu. Rád také přijmu všechny podněty a doporučení k jeho obsahu i zpracování.

Brno, říjen 2018 Autor

1. STATISTIKA A JEJÍ VÝZNAM

V současném světě a zejména v ekonomice má statistika velmi významné a nezastupitelné místo. Moderní řízení ekonomiky v zájmu maximalizace její efektivnosti je nerealizovatelné bez kvalitní informační soustavy a je založeno na neustálém vyhodnocování informací o objektu i jeho okolí za použití exaktních metod. Mimořádně významná role v tomto procesu přísluší statistice, která poskytuje soustavu číselných informací o hospodářství nejenom jako celku, ale také o jeho subsystémech a prvcích. Pro úspěšné použití statistiky je nutné znát cíle, metody a její možnosti a správně interpretovat zjištěné výsledky. Významnou roli v současném rozvoji a využití statistiky hraje výpočetní technika používaná při sběru, přenosu, ukládání a zpracování informací. Počítače se statistickými a databázovými softwarovými pakety (např. systém Statistica, Statgraphics, S-Plus, QC.Expert, Excel aj.) poskytují rozsáhlé možnosti používáni statistických metod.

Slovo ***statistika*** lze chápat ve třech pojetích: jako číselné údaje o hromadných jevech, dále jako praktickou činnost spočívající ve sběru, zpracování a vyhodnocování statistických údajů a jako teoretickou disciplínu, která se zabývá metodami pro popis a odhalování zákonitostí při působení podstatných, relativně stálých činitelů na ***hromadné jevy***, tj. jevy vyskytující se ve velkém měřítku u velkého počtu jedinců (prvků), nazývaných ***statistické jednotky***.

U statistických jednotek sledujeme parametry, charakteristiky, veličiny, ukazatele, indikátory - tzv. ***statistické znaky***. Znaky jsou ***kvantitativní*** (číselné: rozměry, počty apod.) a ***kvalitativní*** (slovní nebo znakové: druhy, třídy apod.). Z informací o nich (naměřených či pozorovaných hodnot), určených vhodným (obvykle náhodným) ***výběrem*** s omezeným ***rozsahem***, získáváme ***statistické soubory***. Tyto soubory pak umožňují popis sledovaných jevů a následné vyvození závěrů pro jevy a procesy. Vzhledem k náhodnosti těchto jevů a ohraničenosti souborů nelze však vyvodit úplné závěry, ale dostatečně blízké sledované realitě.

Na volbě statistických jednotek a vhodném stanovení statistických znaků, pomocí nichž chceme sledovat vlastnosti statistického souboru, závisí úspěch a výsledky veškeré další práce. Proto je třeba správnému vymezení statistického souboru, volbě statistické jednotky a statistických znaků věnovat náležitou pozornost. Statistická jednotka i zjišťované znaky musí být přesně vymezeny z hlediska věcného, které spočívá v přesné definici obsahu zkoumaného znaku.

Statistické zkoumání lze rozdělit do tří etap:

(1) ***statistické zjišťování*** (***šetření***),
(2) ***statistické zpracování***,
(3) ***statistické vyhodnocování a rozbor***.

Etapa (1): Pomocí statistického zjišťování získáváme ***statistické údaje***, což jsou číselné nebo slovní hodnoty (obměny) sledovaných statistických znaků. Nejdříve je nutné určit informační systém (kdo, kdy a jakým způsobem bude zjišťování provádět), který poskytne potřebné informace o statistických jednotkách. Zjišťované údaje mohou být dvojího druhu. Buď se získávají za určitý časový interval (objem produkce, těžba uhlí), nebo jsou vztaženy k určitému okamžiku (stav zásob, počet pracovníků). Pro zjišťování údajů prvního druhu musí být stanovena rozhodná doba (např. objem produkce za rok), pro získávání údajů druhého typu pak rozhodný okamžik (např. počet pracovníků k jistému dni). Dále musí být stanovena doba, rozsah a způsob získávání a zaznamenávání údajů.

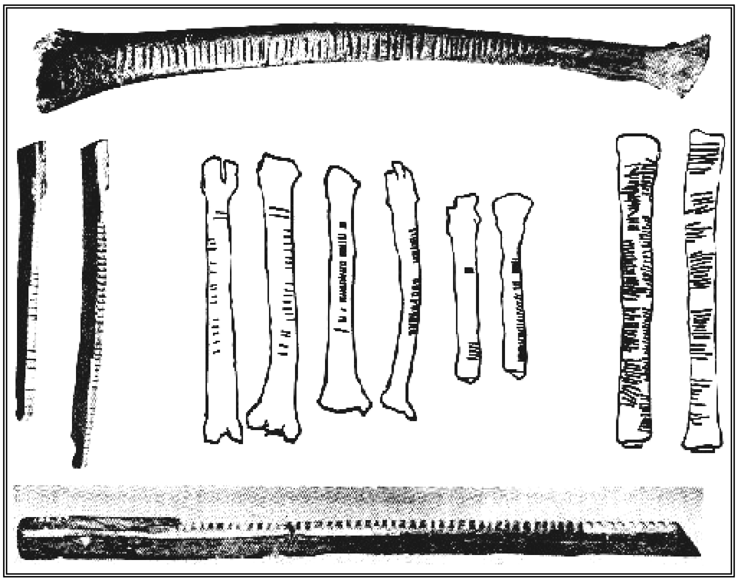
Etapa (2): Výsledkem statistického šetření je velké množství údajů, které je třeba po kontrole a verifikaci určitým způsobem ***utřídit*** a ***shrnout***. Součástí statistického zpracování je proto tabelování a třídění číselných i slovních výsledků, výpočet různých statistických charakteristik, grafické znázorňování výsledných údajů apod. Toto zpracování se v současné době provádí převážně pomocí profesionálního softwaru na počítačích, a to často prostřednictvím lokálních počítačových sítí.

Etapa (3): V závěrečné etapě se provádí ***vyhodnocování*** a ***rozbor*** získaných statistických údajů pomocí vhodných statistických metod. Nezbytnou podmínkou jejich úspěšnosti však je, aby byly správně provedeny etapy předchozí.

Na statistické zkoumání vybraných jevů a procesů pak navazuje ***aplikace výsledků v daném oboru*** (např. změna řízení finančních toků pomocí úrokové sazby a podmínek). Efektivnost aplikace výsledků je však nutno znovu podrobit novému statistickému zkoumání (jde o tzv. ***zpětnou vazbu***).

Statistické metody zpracování a analýzy získaných dat vycházejí z ***popisné statistiky*** (získávání údajů, číselné a grafické zpracování datových souborů), ***teorie pravděpodobnosti*** (náhodné jevy, pravděpodobnost, náhodné veličiny, náhodné procesy a jejich charakteristiky) a ***matematické (inferenční, indukční) statistiky*** (náhodný výběr, odhady parametrů, testy hypotéz, regresní analýza aj.).

Historie těchto disciplín je zajímavá a dosti rozmanitá. Počátky popisné statistiky souvisejí s existencí státních útvarů (jak se tvrdí v úvodech učebnic statistiky) a spadají do období asi 7 tisíc let nazpět. Oprávněně se však lze domnívat, že sahají v čase ještě hlouběji, neboť první záznamy číselného charakteru, vyjadřující spíše množství něčeho než pouhý artefakt, vytvořil pravěký člověk již před 30 tisíci lety záznamy na kostech zvaných "vrubovky“ – viz obr. 1.1.



Obr. 1.1

Základy teorie pravděpodobnosti byly položeny zhruba v XVII. století a dá se říci, že na společenskou objednávku – konkrétně pro řešení nepříliš ušlechtilých, ale zajímavých otázek spojených s požadavky hráčů na dosažení úspěchů při provozování různých víceméně hazardních her. Náznaky stochastického uvažování se však objevují již ve starověku, jak dokládají písemné i obrazové záznamy popisu počtu možných výsledků při hře s hracími kostkami. S řešením konkrétních úloh pomocí teorie pravděpodobnosti v XVII. století jsou spojeny počátky matematické statistiky. Tato disciplína se vyvíjela pozvolně, často s mírným opožděním oproti teorii pravděpodobnosti, a to v souvislosti s nástupem a rozvojem exaktních metod ve vědeckém bádání při studiu a měření reálných jevů i dějů a následnými aplikacemi těchto metod. Její dynamický a expandující vývoj od konce XIX. století byl vyvolán nejen vědeckým, ale především technickým, průmyslovým a ekonomickým rozvojem, tedy potřebami praxe.

Ve zjednodušeném pohledu představují metody matematické statistiky spojení metod popisné statistiky s teorií pravděpodobnosti, a to v tom smyslu, že sledované jevy a procesy studujeme a popisujeme s ohledem na jejich náhodné (stochastické) chování. Přitom náhoda nemusí být obsažena jen v podstatě těchto jevů či procesů (to je v zásadě otázka nazírání a spíše otázka filozofická), ale také v samotném způsobu pozorování nějakého celku pouze prostřednictvím jeho části, která jej sice dostatečně reprezentuje, avšak je víceméně náhodně z celku vybrána.

Ani jedna z uvedených tří disciplín není v žádném případě uzavřenou matematickou disciplínou a dále se rozvíjí. V současné době navíc roste potřeba i možnosti jejich užití velmi dynamicky v souvislosti s nasazením počítačů a využívání datových sítí, a to spolu s rozvojem lidského poznání a konání vyvolává potřebu vývoje nových statistických metod. Jde nejenom o vlastní metody zpracování dat, ale také o způsoby získávání („dolování“) podstatných a relevantních informací ze současných rozsáhlých informačních souborů. Na druhé straně není vždy na místě výsledky získané užitím matematické statistiky přeceňovat, protože např. prokázaná korelace mezi statistickými znaky ještě neznamená jejich kauzalitu.

Kontrolní otázky

1. Jaký význam a postavení má statistika v ekonomice?
2. Jaké tři základní významy má slovo statistika?
3. Co se rozumí statistickým šetřením?
4. Jak probíhá statistické zpracování a vyhodnocování získaných údajů?
5. Z jakých disciplin vycházejí statistické metody?
6. Popište statistické činnosti na konkrétní firmě nebo instituci.

2. POPISNÁ STATISTIKA

Základní pojmy

Při statistickém zkoumání se zabýváme jevy a procesy, které mají hromadný charakter a vyskytují se u rozsáhlého souboru individuálních objektů (výrobky, osoby apod.), nazývaného ***základní soubor*** nebo také ***populace***. Zkoumané objekty jsou tzv. ***statistické jednotky*** a sledujeme u nich vytypované vlastnosti - ***statistické znaky*** (veličiny, parametry atd.), které nabývají pozorovatelných ***hodnot*** (***úrovní***).

Podle druhu hodnot dělíme statistické znaky na ***kvantitativní***, které nabývají číselných hodnot (hmotnost, délka, pevnost, cena, doba, životnost, ...) a ***kvalitativní***, které nemají číselný charakter a lze je vyjádřit slovně (barva, jakostní třída, podmínky provozu, tvar, ...). Sledujeme-li jen jeden znak, hovoříme o ***jednorozměrném*** znaku, naopak o ***vícerozměrném*** znaku.

Kvantitativní znaky dělíme na ***diskrétní***, jestliže nabývají pouze oddělených číselných hodnot (počet zmetků, počet vad, kusová produkce apod.) a ***spojité***, které nabývají všech hodnot z nějakého intervalu reálných čísel (rozměr výrobku, doba do poruchy, cenový index apod.).

Kvalitativní znaky dělíme na ***ordinální***, jejichž slovní hodnoty má smysl uspořádat (jakostní třídy, klasifikace apod.) a ***nominální***, jejichž slovní hodnoty postrádají význam pořadí (barva, tvar, dodavatelé apod.).

Podstatou statistických metod je, že informace o základním souboru nezjišťujeme u všech jeho jednotek, ale jen u některých, které získáme tzv. ***výběrem***. Vedou nás k tomu různá omezení, např. dosažitelnost všech jednotek, velký rozsah základního souboru, způsob získávání informací (zkoušky životnosti, ověření opotřebení atd.), náklady na statistické sledování a další. Počet vybraných jednotek je ***rozsah*** výběru. Dle rozsahu dělíme výběry na ***malé*** (obvykle do 30 až 50) a ***velké*** (řádově stovky, tisíce i více). Toto dělení je relativní a závisí na okolnostech statistického sledování. Výběr by měl být ***reprezentativní*** (poskytovat informace bez omezení) a ***homogenní*** (bez vlivu dalších různých faktorů). To však často nelze v plné míře verifikovatelně zajistit a proto obvykle vybíráme statistické jednotky do výběru ***náhodně***, ovšem s rizikem, že výběr může poskytnout více či méně zkreslené informace o základním souboru. Podle způsobu provedení rozlišujeme výběry:

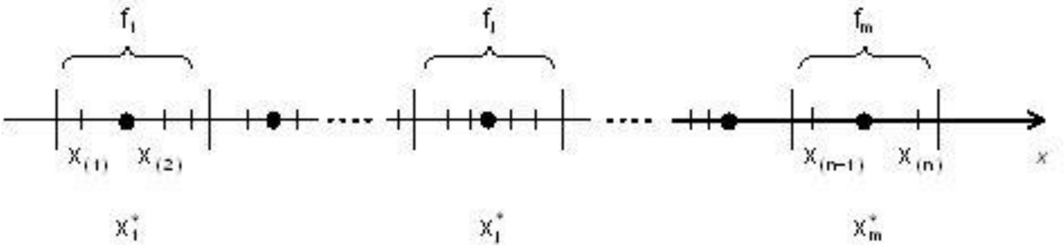
- ***bez opakování*** (každá jednotka může být vybrána nejvýše jednou),
- ***s opakováním*** (každá jednotka může být vybrána vícekrát),
- ***záměrný*** (vybíráme typické jednotky),
- ***oblastní*** (základní soubor rozdělíme na podmnožiny a z nich provedeme části výběru),
- ***systematický*** nebo ***mechanický*** (vybíráme vždy několikátou jednotku co do pořadí při realizaci výběru).

Hodnoty znaku, pozorované či zjištěné na statistických jednotkách z výběru o rozsahu n, tvoří ***statistický soubor s rozsahem*** n. Pro jednorozměrný znak X získáme ***jednorozměrný statistický soubor*** $(x_1,\dots,x_n)$, kde x_i je ***pozorovaná hodnota*** znaku X u i–té statistické jednotky, $i = 1,\dots, n$. Analogicky pro dvourozměrný znak (X, Y) obdržíme ***dvourozměrný statistický soubor*** $((x_1, y_1),\dots,(x_n, y_n))$ apod.

Jednorozměrný statistický soubor s kvantitativním znakem

Získaný statistický soubor $(x_1,\dots,x_n)$ s rozsahem n se také nazývá ***neroztříděný statistický soubor***. Dle potřeby jej můžeme uspořádat podle rostoucích hodnot x_i a obdržíme ***uspořádaný statistický soubor*** $(x_{(1)},\dots,x_{(n)})$, kde $x_{(i)} \le x_{(i+1)}$ pro všechny indexy i. Interval $\left\langle x_{(1)}; x_{(n)} \right\rangle$ je ***variační obor*** a jeho délka $x_{(n)} - x_{(1)}$ je ***rozpětí*** statistického souboru.

Při velkém rozsahu statistického souboru nebo z důvodu dalšího zpracování (některá grafická vyjádření anebo užití matematicko - statistických metod) původní soubor ***roztřídíme***. ***Roztříděný statistický soubor*** (viz obr. 2.1) získáme pokrytím variačního oboru systémem disjunktních intervalů (obvykle zleva otevřených a zprava uzavřených), tzv. ***tříd*** o počtu *m*, které mají obvykle stejnou ***délku*** *h*. Každá třída je reprezentována uspořádanou dvojicí (x_j^*, f_j), kde x_j^* je ***střed*** *j*-té třídy, $x_j^* < x_{j+1}^*$, a f_j je ***absolutní četnost*** *j*-té třídy, $j = 1,...,m$. Absolutní četnost f_j je počet prvků x_i původního neroztříděného statistického souboru, které leží v *j*-té třídě. Číslo $\frac{f_j}{n}$ je ***relativní četnost*** a uvádí se též v %. Platí $\sum_{j=1}^{m} f_j = n$.



Obr. 2.1

Počet tříd *m* volíme obvykle přibližně $1 + 3{,}3\log n$ (pro statistický soubor symetrického charakteru) anebo $\sqrt{n}$ až $2\sqrt{n}$ (pro statistický soubor asymetrického charakteru). ***Délka*** třídy je $h \approx \frac{x_{(n)} - x_{(1)}}{m}$ a stanovujeme ji tak, aby odpovídala přesnosti získání hodnot x_i a aby střed třídy x_j^* byl zaokrouhlené číslo. U diskrétního znaku volíme obvykle za středy tříd přímo hodnoty, kterých tento znak může nabývat. Pokud třídění provádíme na PC, měli bychom zkontrolovat, zda nastavení parametrů *m*, resp. *h* použitého statistického software odpovídá našim požadavkům.

Číslo $F_j = \sum_{k=1}^{j} f_k$ je ***kumulativní absolutní četnost***, číslo $\frac{F_j}{n}$ je ***kumulativní relativní četnost***, $j = 1,...,m$, a uvádí se též v %. Platí, že $F_{j+1} = F_j + f_{j+1}$ pro $j = 1,...,m-1$, kde $F_1 = f_1$, takže $F_m = n$.

Roztříděný statistický soubor zapisujeme do tzv. ***četnostní tabulky*** pro různé typy četností, např. pro absolutní četnosti:

x_j^*	x_1^*	...	x_m^*
f_j	f_1	...	f_m

Významné vlastnosti statistického souboru vyjadřují v koncentrované formě jeho následující ***číselné*** (***empirické***) ***charakteristiky***. Jde zejména o ***charakteristiky polohy, proměnlivosti a souměrnosti***.

Základní ***charakteristiky polohy*** statistického souboru jsou:

1. ***Aritmetický průměr***

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{pro neroztříděný soubor,}$$

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{m} f_j x_j^* \qquad \text{pro roztříděný soubor.}$$

Vlastnosti aritmetického průměru jsou:

a) $y = ax + b \Rightarrow \bar{y} = a\bar{x} + b$ pro reálné konstanty *a*, *b*,

b) $\overline{x+y} = \bar{x} + \bar{y}$ pro soubory se stejným rozsahem,

c) $x_{(1)} \le \bar{x} \le x_{(n)}$,

d) $\bar{x}$ má tentýž rozměr jako znak *X*.

Někdy se užívá též ***vážený aritmetický průměr***

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i},$$

kde $w_i \ge 0$ jsou ***váhy*** (vhodně stanovená reálná čísla, z nichž aspoň jedno je nenulové) hodnot x_i, které vyjadřují jejich význam, např. přesnost.

2. ***Medián*** neroztříděného statistického souboru

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{pro lichá } n\,, \\ \frac{1}{2}\left[x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right] & \text{pro sudá } n\,. \end{cases}$$

Vlastnosti mediánu:

a) $y = ax + b \Rightarrow \tilde{y} = a\tilde{x} + b$ pro reálné konstanty *a*, *b*,

b) $(x + y)^{\sim} = \tilde{x} + \tilde{y}$ pro uspořádané soubory se stejným rozsahem,

c) $x_{(1)} \le \tilde{x} \le x_{(n)}$,

d) $\tilde{x}$ má tentýž rozměr jako znak *X*.

Medián rozděluje statistický soubor na "dolní polovinu" a "horní polovinu" hodnot x_i (viz obr. 2.1). Jde o ***robustní*** charakteristiku, která je oproti aritmetickému průměru málo citlivá na extrémně odchýlené hodnoty. Pro roztříděný soubor se k výpočtu mediánu užívá vhodná aproximace.

3. ***Modus*** $\hat{x}$ je číslo, v jehož okolí je nejvíce hodnot x_i, resp. je to střed x_j^* třídy s největší absolutní četností f_j. Modus má tytéž vlastnosti jako aritmetický průměr i medián a dle potřeby se počítá vhodnou aproximací pro roztříděný soubor).

Základní ***charakteristiky proměnlivosti*** (***variability***) statistického souboru jsou:

1. ***Rozptyl*** (***disperze***, ***variance***)

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2 \qquad \text{pro neroztříděný soubor,}$$

$$s^2 = \frac{1}{n}\sum_{j=1}^{m} f_j (x_j^* - \bar{x})^2 = \left(\frac{1}{n}\sum_{j=1}^{m} f_j x_j^{*2}\right) - \bar{x}^2 \qquad \text{pro roztříděný soubor.}$$

Dle potřeby a také pro zdůraznění znaku *X* někdy píšeme $s^2(x)$ apod. Vlastnosti rozptylu jsou:

a) $s^2 \geq 0$,

b) $y = ax + b \Rightarrow s^2(y) = a^2 s^2(x)$ pro reálné konstanty *a*, *b*,

c) $s^2 = 0 \Leftrightarrow x_1 = \cdots = x_n$, resp. $x_1^* = \cdots = x_m^*$,

d) s^2 má rozměr rovný kvadrátu rozměru znaku *X*.

Větší proměnlivosti znaku *X* odpovídá větší rozptyl a naopak. Při výpočtech se také užívá jiný vzorec pro rozptyl, když výraz $\frac{1}{n}$ zaměníme výrazem $\frac{1}{n-1}$. Takto vypočtený rozptyl je roven číslu $\frac{n}{n-1}s^2 > s^2$ (pro $s^2 \neq 0$). Zdůvodnění výrazu $\frac{1}{n-1}$ plyne z požadavků tzv. nestranného odhadu rozptylu základního souboru.

2. ***Směrodatná odchylka*** $s = \sqrt{s^2}$.

Dle potřeby také píšeme *s*(*x*). Vlastnosti směrodatné odchylky jsou:

a) $s \geq 0$,

b) $y = ax + b \Rightarrow s(y) = |a|\, s(x)$ pro reálné konstanty *a*, *b*,

c) $s = 0 \Leftrightarrow x_1 = \cdots = x_n$, resp. $x_1^* = \cdots = x_m^*$

d) s má tentýž rozměr jako znak X .

Větší proměnlivosti znaku *X* odpovídá větší směrodatná odchylka a naopak.

3. ***Variační koeficient*** $v = \frac{s}{\bar{x}}$.

Dle potřeby také píšeme *v*(*x*). Vlastnosti variačního koeficientu jsou:

a) $v(ax) = \frac{a}{|a|} v(x)$ pro reálnou konstantu $a \neq 0$,

b) v je bezrozměrné číslo.

Jde o relativní míru variability znaku *X* a uvádí se též v %. Má smysl pouze pro znak *X*, který nabývá pouze kladných anebo záporných hodnot. Není proto např. vhodný pro znak *X* vyjadřující odchylky od nějaké nominální hodnoty.

4. ***Rozpětí*** $x_{(n)} - x_{(1)}$. Rozpětí má stejné vlastnosti jako směrodatná odchylka.

Základní ***charakteristikou souměrnosti*** statistického souboru je ***koeficient šikmosti*** (***koeficient asymetrie***)

$$A = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} \qquad \text{pro neroztříděný soubor,}$$

$$A = \frac{\frac{1}{n}\sum_{j=1}^{m} f_j (x_j^* - \bar{x})^3}{s^3} \qquad \text{pro roztříděný soubor.}$$

Dle potřeby také píšeme *A*(*x*). Vlastnosti koeficientu šikmosti jsou:

a) $A > 0 \Leftrightarrow$ většina hodnot x_i je menší než (leží pod) $\bar{x}$,

b) $A = 0 \Leftrightarrow$ hodnoty x_i jsou rozloženy souměrně vzhledem k $\bar{x}$,

c) $A < 0 \Leftrightarrow$ většina hodnot x_i je větší než (leží nad) $\bar{x}$,

d) $y = ax + b \Rightarrow A(y) = \dfrac{a}{|a|} A(x)$ pro reálné konstanty a, b, $a \neq 0$,

e) A je bezrozměrné číslo.

Existuje řada dalších číselných charakteristik statistického souboru. Např. pro poměrové znaky (cenové a objemové indexy, úrokové míry apod.) se místo aritmetického průměru užívá ***geometrický průměr***

$$\bar{x}_g = \sqrt[n]{x_1 \ldots x_n}$$

a ve speciálních případech (např. pro znaky vyjadřující rychlost nějakého děje) počítáme ***harmonický průměr***

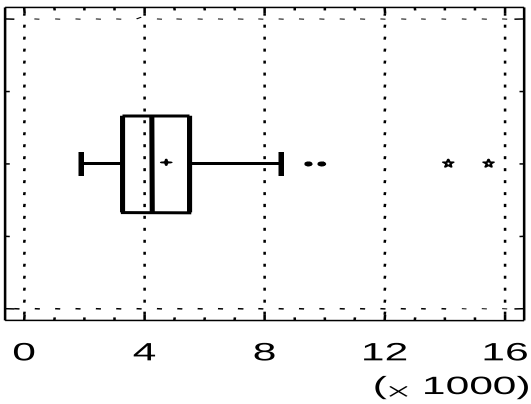
$$\bar{x}_h = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1}.$$

Dle potřeby se také někdy používá ***koeficient špičatosti*** (***koeficient excesu***)

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3,$$

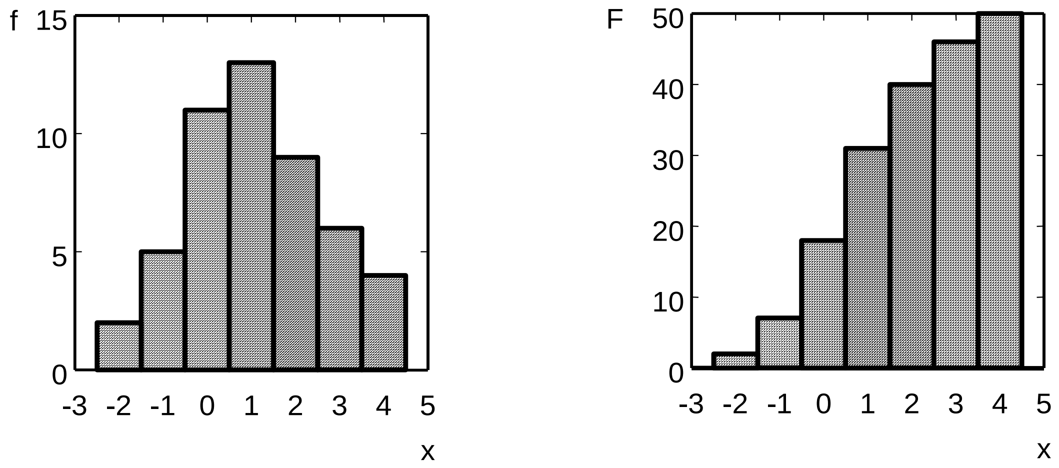
který vyjadřuje specifickým způsobem míru koncentrace hodnot statistického souboru.

Mnoho rychlých a cenných informací poskytují o statistických souborech jejich ***grafická vyjádření***. Pro jednorozměrný neroztříděný resp. uspořádaný statistický soubor se zejména užívá ***krabicový graf*** - obr. 2.2, kde obdélník obsahuje střední část uspořádaného souboru (cca polovinu všech prvků) tak, že nalevo a napravo od něj leží vždy cca čtvrtina jeho prvků. Levá (pravá) svislá čára odpovídá tzv. ***dolnímu*** (***hornímu***) ***kvartilu*** a svislá čára uvnitř je v místě ***mediánu***. Výška obdélníku je úměrná rozsahu souboru a úsečky ("vousy") vlevo a vpravo vyjadřují přijatelné obory pro uvedené čtvrtiny souboru. Prvky mimo tyto úsečky jsou považovány za podezřelé, případně extrémně odchýlené. Existují další modifikace tohoto grafu a jiná vyjádření.

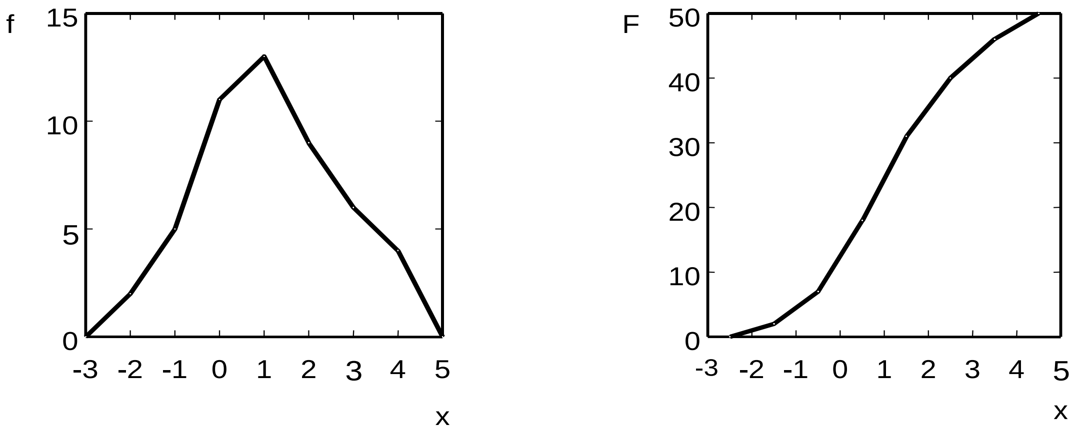


Obr. 2.2

Pro jednorozměrný roztříděný statistický soubor se v případě spojitého znaku užívají nejčastěji následující dva typy grafů. ***Histogram*** - obr. 2.3 - je soustava obdélníků v kartézské souřadné soustavě, jejichž základny jsou třídy a výšky jsou četnosti tříd (absolutní, relativní, kumulativní atd.). ***Polygon*** - obr. 2.4 - je lomená čára v kartézské souřadné soustavě spojující body, jejichž x-ová souřadnice je střed (příp. horní hranice) třídy a y-ová souřadnice je četnost třídy.

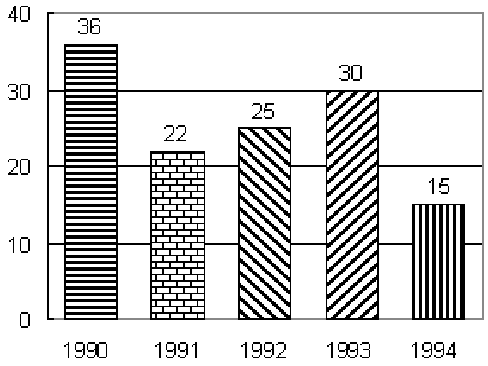


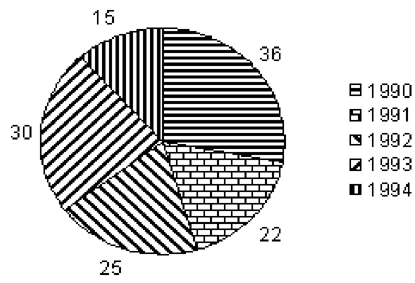
Obr. 2.3



Obr. 2.4

Pro jednorozměrný roztříděný statistický soubor s diskrétním znakem se užívají obvykle následující grafy. ***Sloupcový graf*** - obr. 2.5 - je podobný histogramu, avšak obdélníky na sebe nenavazují a někdy se kreslí ve vodorovné poloze. ***Výsečový*** (***koláčový***) ***graf*** - obr. 2.6 - je kruh rozdělený na výseče, jejichž vnější obvod odpovídá četnostem tříd, případně jsou zvolené výseče vysunuty. V uvedených grafech se různými barvami nebo šrafováním zvýrazňují potřebné informace a mnohdy se dále geometricky a výtvarně prezentačně modifikují.





Obr. 2.5 Obr. 2.6

Příklad 2.1

Měřením délky X (mm) 10 válečků byly získány hodnoty: 5,38; 5,36; 5,35; 5,40; 5,41; 5,34; 5,29; 5,43; 5,42; 5,32. Určete rozsah, variační obor, variační rozpětí, aritmetický průměr $\bar{x}$, rozptyl s^2, směrodatnou odchylku s, variační koeficient v a medián $\tilde{x}$ daného statistického souboru.

Ř e š e n í:

Rozsah daného souboru je $n = 10$, takže nemá smysl jej třídit. Protože $x_{(1)} = 5{,}29$ mm a $x_{(10)} = 5{,}43$ mm, je variační obor $\langle 5{,}29;\ 5{,}43 \rangle$ mm a variační rozpětí je 5,43 – 5,29 = 0,14 mm. Dále je:

$\bar{x} = (5{,}38 + \ldots + 5{,}32)/10 = 53{,}70/10 = 5{,}37$ mm,

$s^2 = (5{,}38^2 + \ldots + 5{,}32^2)/10 - 5{,}37^2 = 288{,}388/10 - 28{,}8369 = 0{,}0019$ mm^2,

$s = \sqrt{0{,}0019} \doteq 0{,}0435889894 \doteq 0{,}044$ mm,

$v = \sqrt{0{,}0019}/5{,}37 \doteq 0{,}0435889894/5{,}37 \doteq 0{,}00811713 \doteq 0{,}8117\ \%$,

$\tilde{x} = (5{,}36 + 5{,}38)/2 = 5{,}37$ mm.

Modus nemá smysl určovat. Pro grafické vyjádření souboru by byl vhodný krabicový graf.

Příklad 2.2

Při kontrole byl zjišťován objem stejně baleného nápoje v 50 lahvích a byly naměřeny následující odchylky X (ml) od hodnoty na etiketě:

1,2;	2,1;	1,7;	0,9;	0,3;	2,0;	-1,3;	-0,1;	3,2;	2,8;
0,8;	4,4;	2,9;	1,2;	0,0;	-2,3;	1,2;	0,9;	2,3;	-0,2;
0,1;	1,9;	-1,9;	-0,2;	-1,3;	1,5;	0,5;	2,0;	-1,3;	3,7;
0,9;	1,0;	0,4;	1,9;	1,4;	-1,3;	1,6;	1,4;	3,1;	-0,1;
1,8;	0,0;	4,1;	1,3;	3,0;	0,4;	3,8;	-0,8;	3,1;	0,9.

Roztřiďte daný statistický soubor, graficky jej znázorněte a vypočtěte $\bar{x}$, s^2, s, $\hat{x}$, A.

Ř e š e n í:

Rozsah souboru $n = 50$; $x_{(1)} = -2{,}3$ ml a $x_{(50)} = 4{,}4$ ml, takže variační obor je $\langle -2{,}3;\ 4{,}4 \rangle$ ml a rozpětí je 4,4 – (–2,3) = 6,7 ml. Volíme počet tříd $m = 7$ (tj. asi $\sqrt{50}$) a délku třídy $h = 1$ (tj. asi 6,7/7). Volba tříd a jejich středů, roztřídění do tříd a výpočet absolutních a kumulativních četností je v následující tabulce, kde např. // značí 2 hodnoty a ~~////~~ značí 5 hodnot ležících v dané třídě:

j	třída	x_j^*	zařazení do tříd	f_j	F_j
1	-2,5; -1,5	-2	//	2	2
2	-1,5; -0,5	-1	~~////~~	5	7
3	-0,5; 0,5	0	~~////~~ ~~////~~ /	11	18
4	0,5; 1,5	1	~~////~~ ~~////~~ ///	13	31
5	1,5; 2,5	2	~~////~~ ////	9	40
6	2,5; 3,5	3	~~////~~ /	6	46
7	3,5; 4,5	4	////	4	50

Histogramy a polygony tohoto statistického souboru jsou na obr. 2.3 a 2.4. Další výpočty jsou pro přehlednost znázorněny v následující tabulce, ze které dostaneme:
$\overline{x}$ = 56/50 = 1,12 ml;
s^2 = 180/50 − $1{,}12^2$ = 2,3456 ml^2;
$s = \sqrt{2{,}3456} \doteq$ 1,532 ml;
střed třídy s největší četností $\hat{x}$ = 1 ml;
dalším výpočtem obdržíme $A \doteq$ 0,098502.

j	x_j^*	f_j	$f_j x_j^*$	$f_j x_j^{*2}$
1	-2	2	-4	8
2	-1	5	-5	5
3	0	11	0	0
4	1	13	13	13
5	2	9	18	36
6	3	6	18	54
7	4	4	16	64
$\sum$	—	50	56	180

Dvourozměrný statistický soubor s kvantitavními znaky

Pozorovaný statistický soubor $((x_1, y_1),...,(x_n, y_n))$ s rozsahem n je ***neroztříděný statistický soubor***. Vynecháním první, resp. druhé, hodnoty v každé dvojici obdržíme jednorozměrné statistické soubory $(x_1,...,x_n)$ a $(y_1,...,y_n)$. Zpracováním těchto souborů získáme jejich číselné charakteristiky $\overline{x}$, $\overline{y}$, $s^2(x)$, $s^2(y)$ atd.

Roztříděný dvourozměrný statistický soubor získáme roztříděním jednorozměrných statistických souborů $(x_1,...,x_n)$ a $(y_1,...,y_n)$, přičemž oba roztříděné soubory mohou mít různé počty tříd i jejich délky. Dostaneme tak dvourozměrné třídy se ***středy*** (x_j^*, y_k^*) a ***absolutními četnostmi*** f_{jk}, $j = 1,...,m_1$ a $k = 1,...,m_2$. Dle potřeby se dále určují ***relativní četnosti*** $\frac{f_{jk}}{n}$, ***kumulativní četnosti*** F_{jk} atd.

Roztříděný dvourozměrný statistický soubor zapisujeme do ***četnostní tabulky*** pro různé typy četností. Následující tabulka je pro absolutní četnosti f_{jk}, kde čísla f_{xj} a f_{yk} jsou ***marginální*** (***okrajové***) ***četnosti*** a platí

$$f_{xj} = \sum_{k=1}^{m_2} f_{jk}\ ,\quad f_{yk} = \sum_{j=1}^{m_1} f_{jk}\ ,\quad \sum_{j=1}^{m_1} f_{xj} = \sum_{k=1}^{m_2} f_{yk} = \sum_{j=1}^{m_1}\sum_{k=1}^{m_2} f_{jk} = n\,.$$

$\diagdown \; y_k^* \;/\; x_j^*$	y_1^*	...	$y_{m_2}^*$	f_{xj}
x_1^*	f_{11}	...	$f_{1\,m_2}$	$f_{x\,1}$
...	...	...	...	...
$x_{m_1}^*$	$f_{m_1 1}$	...	$f_{m_1\,m_2}$	$f_{x\,m_1}$
f_{yk}	$f_{y\,1}$	...	$f_{y\,m_2}$	n

Pro roztříděné jednorozměrné statistické soubory $\left(x_j^*, f_{xj}\right)$, $j = 1,\dots,m_1$, a $\left(y_k^*, f_{yk}\right)$, $k = 1,\dots,m_2$, obdržíme jejich číselné charakteristiky $\overline{x}$, $\overline{y}$, $s^2(x)$, $s^2(y)$ atd.

Mírou závislosti znaků *X* a *Y* je ***koeficient korelace*** (***korelační koeficient***)

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{s(x)s(y)} = \frac{\frac{1}{n}\sum_{i=1}^{n}x_i y_i - \overline{x}\overline{y}}{s(x)s(y)} \quad \text{pro neroztříděný soubor,}$$

$$r = \frac{\frac{1}{n}\sum_{j=1}^{m_1}\sum_{k=1}^{m_2} f_{jk}(x_j^* - \overline{x})(y_k^* - \overline{y})}{s(x)s(y)} = \frac{\frac{1}{n}\sum_{j=1}^{m_1}\sum_{k=1}^{m_2} f_{jk}x_j^* y_k^* - \overline{x}\overline{y}}{s(x)s(y)} \quad \text{pro roztříděný soubor,}$$

přičemž čitatelé ve všech zlomcích vyjadřují tzv. ***kovarianci***, kterou značíme *cov*. Někdy pro zdůraznění znaků *X, Y* píšeme *r*(*x, y*), resp. *cov*(*x, y*). Vlastnosti koeficientu korelace:

a) $u = ax + b,\ v = cy + d \Rightarrow r(u,v) = \frac{ac}{|ac|} r(x,y)$ pro reálné konstanty $a, b, c, d,\ a \neq 0, c \neq 0$,

b) $r(y,x) = r(x,y)$,

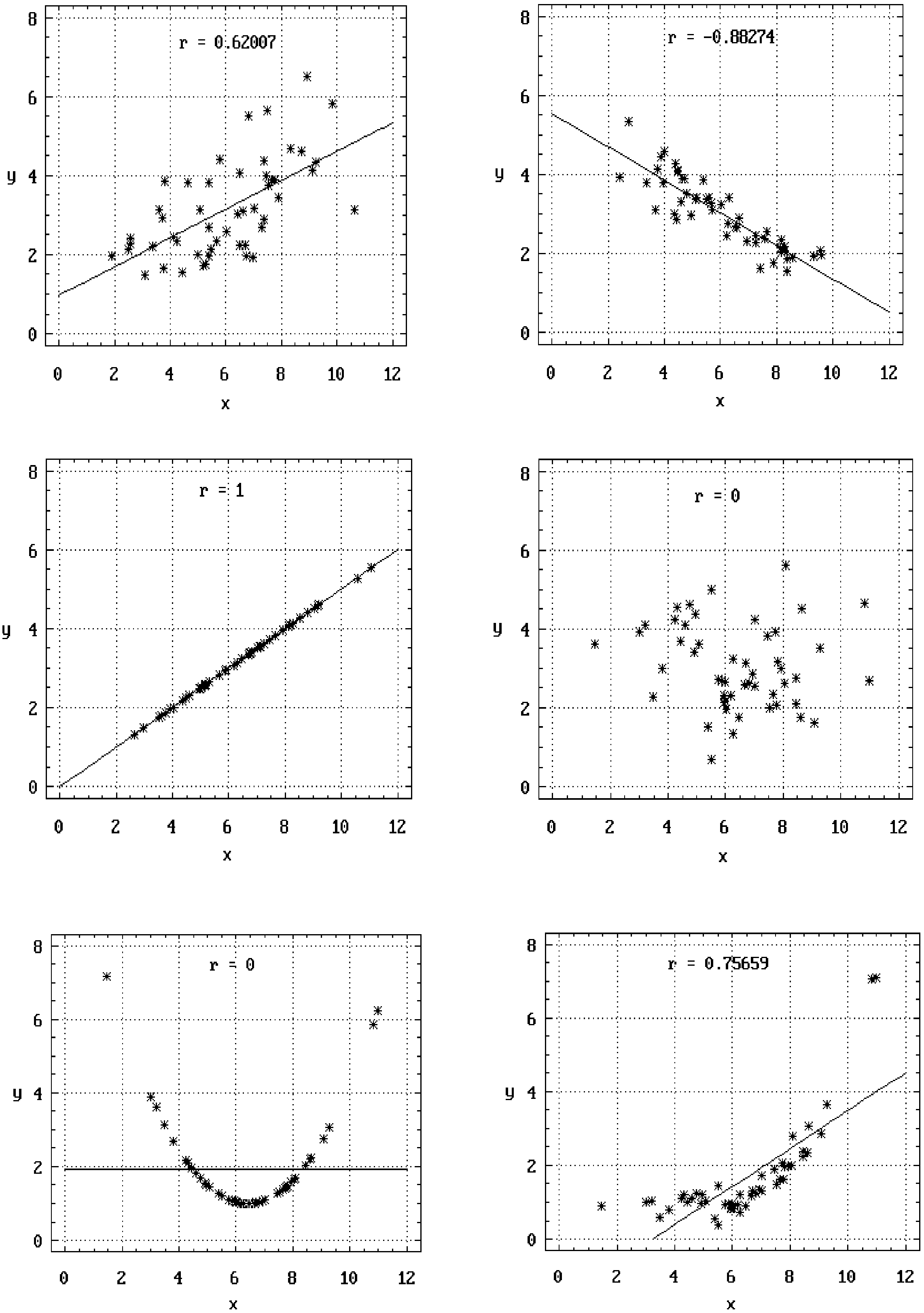
c) $-1 \leq r \leq 1$,

d) $r = \pm 1 \Leftrightarrow y = ax + b, a \neq 0$,

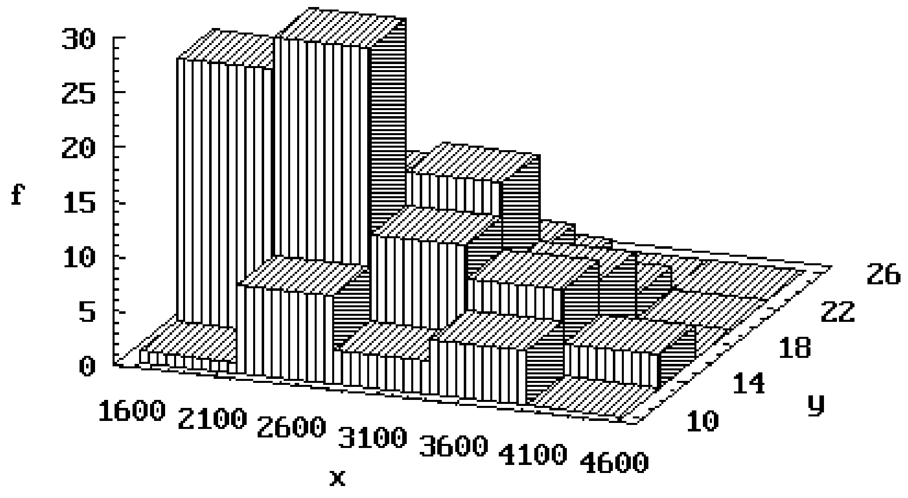
e) *r* je bezrozměrné číslo.

Koeficient korelace *r* je pouze mírou lineární závislosti mezi znaky *X* a *Y*. Čím je jeho hodnota bližší 1 anebo -1, tím je závislost bližžší lineární závislosti a body $\left(x_i, y_i\right)$ bližší přímce. Jeho kladná (záporná) hodnota odpovídá převážně rostoucí (klesající) závislosti mezi *X* a *Y*. Hodnota blízká 0 vyjadřuje, že závislost není lineární a znaky *X, Y* mohou být nezávislé.

Pro grafické vyjadření dvourozměrného neroztříděného statistického souboru se užívá ***rozptylový graf*** na obr. 2.7, kde jsou rovněž uvedeny pro ilustraci hodnoty koeficientu korelace, a pro dvourozměrný roztříděný statistický soubor třírozměrný ***histogram*** na obr. 2.8, případně třírozměrný ***sloupcový graf*** pro diskrétní znaky *X, Y*.



Obr. 2.7



Obr. 2.8

Příklad 2.3
Statistickým šetřením nákladů X (Kč) a cen Y (Kč) pro stejný výrobek u 10 výrobců byl získán dvourozměrný statistický soubor:

(30,18; 50,26), (30,19; 50,23), (30,21; 50,27), (30,22; 50,25), (30,25; 50,22),
(30,26; 50,32), (30,26; 50,33), (30,28; 50,29), (30,30; 50,37), (30,33; 50,42).

Vypočtěte $\bar{x}$, $\bar{y}$, $s^2(x)$, $s^2(y)$, $s(x)$, $s(y)$, c, r.

Ř e š e n í:

Vzhledem k malému rozsahu n = 10 soubor netřídíme. Použitím výše uvedených vztahů dostaneme:

$\bar{x}$ = (30,18 + ... + 30,33)/10 = 30,248 Kč … průměrné náklady,

$\bar{y}$ = (50,26 + ... + 50,42)/10 = 50,296 Kč … průměrná cena,

$s^2(x)$ = (30,18^2 + ... + 30,33^2)/10 - 30,248^2 = 0,002096 Kč^2,

$s^2(y)$ = (50,26^2 + ... + 50,42^2)/10 - 50,296^2 = 0,003684 Kč^2,

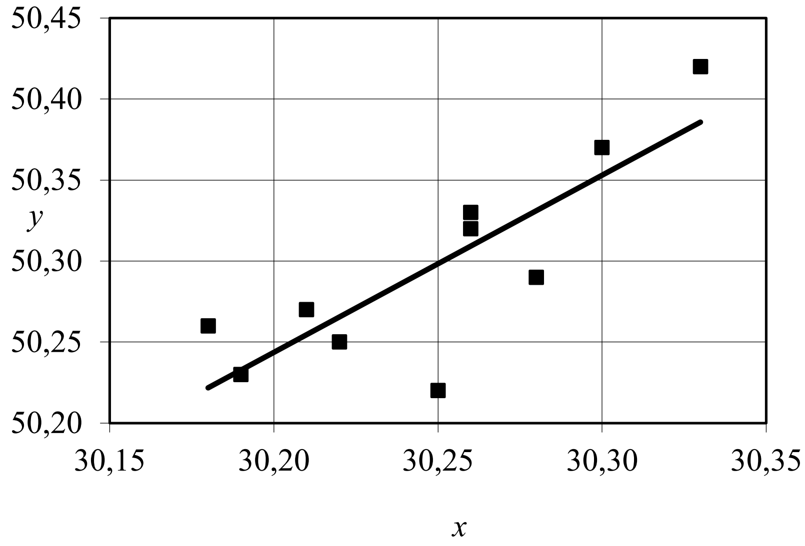
$s(x) = \sqrt{0,002096} \doteq$ 0,0457821 Kč $\doteq$ 0,0458 Kč,

$s(y) = \sqrt{0,003684} \doteq$ 0,0606960 Kč $\doteq$ 0,0607 Kč,

cov = (30,18.50,26 + ... + 30,33.50,42)/10 - 30,248.50,296 = 0,002292 Kč^2,

r = 0,002292/(0,0457821·0,0606960) = 0,82481996263 $\doteq$ 0,8248.

Vzhledem k velikosti koeficientu korelace r lze předpokládat, že mezi oběma znaky X a Y (náklady a cenou) je závislost víceméně blízká lineární. Jeho kladná hodnota odpovídá tomu, že s rostoucími náklady roste cena výrobku. Rozptylový graf daného statistického souboru je na obr. 2.9.



Obr. 2.9

Statistické soubory s kvalitativními znaky

Jednorozměrný statistický soubor s kvalitativním znakem $(x_1,...,x_n)$ s rozsahem n vyjadřujeme pomocí ***četnostní tabulky,*** kde x_j^* jsou možné slovní hodnoty znaku X a f_j jsou četnosti těchto hodnot v původním souboru, $j = 1,...,m$. Číselné charakteristiky se až na výjimky nepoužívají – viz [1], [3] a [4]. Ke grafickému vyjádření souboru slouží ***sloupcový graf***, ***koláčový graf*** apod. ***Dvourozměrný statistický soubor s kvalitativními znaky*** $((x_1, y_1),...,(x_n, y_n))$ s rozsahem n vyjadřujeme pomocí ***četnostní tabulky*** podobně jako pro kvantitativní znaky, kde (x_j^*, y_k^*) jsou dvojice možných slovních hodnot dvourozměrného kvalitativního znaku (X, Y) a f_{jk} jsou četnosti těchto hodnot v původním souboru pro $j = 1,...,m_1$ a $k = 1,...,m_2$. Z číselných charakteristik se užívají především různé míry závislosti znaků X a Y - viz [1], [3] a [4]. Ke grafickému vyjádření souboru slouží třírozměrný ***sloupcový graf*** podobný třírozměrnému sloupcovému grafu pro dvourozměrný diskrétní kvantitativní znak.

Kontrolní otázky

1. Popište typy statistických znaků a uveď te konkrétní příklady.
2. Co je výběr, jaké má vlastnosti a jak jej provádíme?
3. Definujte statistický soubor a uveďte, jak souvisí se základním souborem.
4. Popište roztřídění jednorozměrného statistického souboru s kvantitativním znakem.
5. Uveďte charakteristiky polohy jednorozměrného statistického souboru s kvantitativním znakem, jejich vlastnosti a význam.
6. Uveďte charakteristiky variability a souměrnosti jednorozměrného statistického souboru s kvantitativním znakem, jejich vlastnosti a význam.
7. Popište grafická znázornění jednorozměrného statistického souboru s kvantitativním znakem.
8. Popište roztřídění dvourozměrného statistického souboru s kvantitativními znaky a jeho grafická znázornění.

9. Uveďte číselné charakteristiky dvourozměrného statistického souboru s kvantitativními znaky.
10. Jaké vlastnosti a význam má koeficient korelace?
11. Popište zpracování a grafická znázornění statistických souborů s kvalitativními znaky.
12. Uveďte statistické znaky, které charakterizují ukazatele o Vaší firmě.

3. ANALÝZA ČASOVÝCH ŘAD

Základní pojmy

Většina ekonomické jevů se chová dynamicky, tj. vyvíjí se v čase. Základním prostředkem studia dynamiky takových jevů je analýza jejich vývoje v minulosti, která nám umožňuje poznat existující zákonitosti sledovaných jevů na čase a na základě tohoto poznání předpovídat jejich chování v budoucnosti.

Časovou řadu dostaneme, když údaje o sledovaném jevu ve sledovaném časovém úseku chronologicky uspořádáme. Dobře sestavená a pro analýzu použitelná časová řada musí splňovat tyto požadavky:

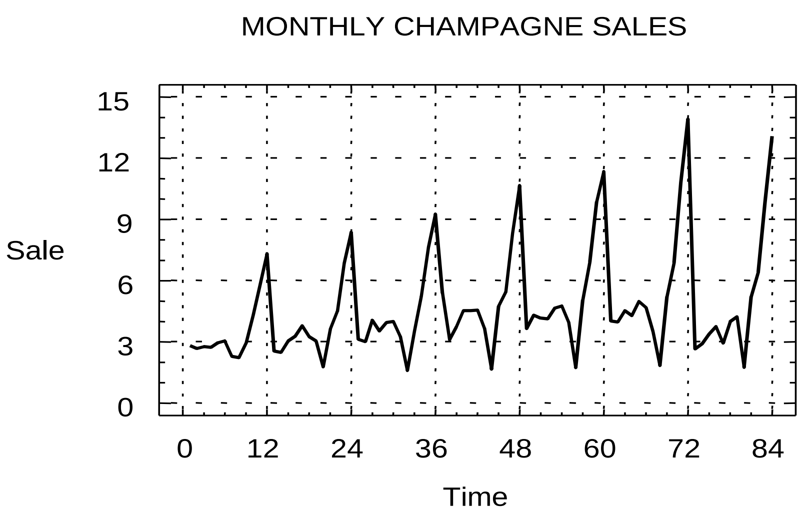
- údaje musí být seřazené chronologicky,
- údaje musí být porovnatelné, jinak řečeno musí být zajištěna:
  a) jednota časového období, ve kterém jsou získány,
  b) jednotná definice údaje (měrné jednotky, stejný způsob sběru dat).

Pokud některé z uvedených podmínek nerespektujeme, získáme nesprávné závěry.

Z hlediska matematické statistiky je ***časová řada*** posloupnost

$$(y_1, \ldots, y_n)$$

pozorovaných hodnot y_i statistického znaku Y, kde index i odpovídá časovému okamžiku t_i nebo i-tému intervalu končícímu v t_i, k němuž se y_i vztahuje, $t_i < t_{i+1}$, $i = 1, 2, \ldots$. Někdy místo y_i píšeme y_t. Graficky se časová řada nejčastěji znázorňuje pomocí ***grafu*** v kartézské souřadné soustavě, kde na x-ovou osu vynášíme indexy i anebo časy t_i a na y-ovou osu hodnoty y_i Příklad grafu časové řady je na obr. 3.1.



Obr. 3.1

Časové řady se vztahují k určitému období, pak se jedná o ***intervalové časové řady***, anebo k určitému okamžiku, kdy se jedná o ***okamžikové časové řady***.

Intervalové časové řady obsahují ukazatele, zjišťované vždy za určité časové období (hodina, den, měsíc, rok, atd.). Pro tyto časové řady je charakteristické, že:

- údaje vyjadřují množství,
- jsou závislé na délce sledovaného časového intervalu,
- součet údajů má určitý význam a smysl.

Okamžikové časové řady obsahují údaje, vztažené k určitému časovému okamžiku. Pro tyto časové řady je charakteristické, že:

- údaje vyjadřují úroveň nebo stav zkoumaného jevu,
- údaje nejsou závislé na době mezi sledovanými časovými okamžiky,
- součet jednotlivých údajů nemá konkrétní smysl.

Pro správný rozbor časové řady je nutné si uvědomit určité rozdíly, které plynou z odlišného charakteru údajů obsažených v časových řadách a jejich významu. Proto rozdělujeme časové řady na řady:

1) ***původních veličin***:	a) intervalové	
	b) okamžikové	
2) ***odvozených veličin***:	a) součtové:	α) kumulativní
		β) klouzavých součtů
	b) průměrové:	α) kumulativních průměrů
		β) klouzavých průměrů
	c) rozdílové a poměrové	

Průměry časových řad

Intervalové řady

Hodnoty intervalových veličin (znaků) pro intervalové časové řady se vztahují k určitým intervalům (časovým obdobím) a tyto hodnoty podstatně ovlivňují délky těchto intervalů. Nejběžnějšími intervalovými veličinami jsou: produkce, maloobchodní obrat, tržby, mzdové fondy, odpracované hodiny, počet narozených dětí v určitém období atd. Jejich základní agregující číselnou charakteristikou je ***aritmetický průměr*** (viz kapitolu 2) hodnot $y_1,\dots,y_n$. Pro stejně dlouhé časové intervaly používáme prostý aritmetický průměr

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i$$

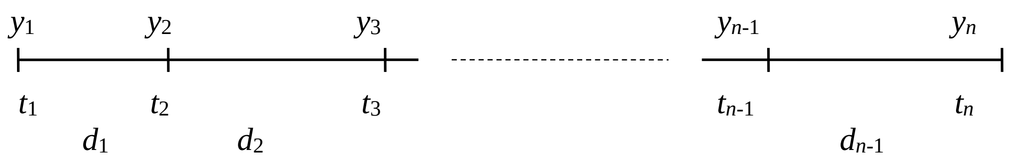
a pro různě dlouhé časové intervaly je vhodný vážený aritmetický průměr s váhami rovnými převráceným hodnotám těchto délek.

Okamžikové řady

Hodnoty okamžikových veličin (znaků) se nevztahují na určité období, ale k určitému okamžiku. Tímto okamžikem může být první nebo poslední den tohoto období, záměrně zvolený den nebo okamžik. Příkladem okamžikové veličiny údaje je např. početní stav obyvatelstva, dělníků, stav základních prostředků atd. Tyto údaje ukazují okamžitý stav zvoleného jevu. Pro jejich základní agregování používáme tzv. ***chronologický průměr***

$$\bar{y}_{chr}=\frac{y_1 d_1+y_2(d_1+d_2)+\dots+y_{n-1}(d_{n-2}+d_{n-1})+y_n d_{n-1}}{2\left(d_1+d_2+\dots+d_{n-1}\right)},$$

kde $d_i=t_{i+1}-t_i,\ i=1,\dots,n-1$, jsou délky intervalů mezi časovými okamžiky t_i, v nichž byly pozorovány hodnoty y_i dané časové řady – viz obr. 3.2.



Obr. 3.2

Při výpočtu chronologického průměru vlastně stanovujeme váhy hodnot y_i tak, že vnitřní hodnota $y_2,\ldots,y_{n-1}$ zastupuje polovinu předcházejícího a polovinu následujícího časového intervalu. Krajním hodnotám y_1, resp. y_n, dáváme jenom váhu poloviny následujícího, resp. předcházejícího, časového intervalu. Speciálně při stejných časových vzdálenostech $d_1 = \ldots = d_{n-1}$ je chronologický průměr

$$\bar{y}_{chr} = \frac{0,5y_1 + y_2 + y_3 + \cdots + y_{n-1} + 0,5y_n}{n-1}.$$

Příklad 3.1

Ve výrobní firmě ZKZS, s.r.o. je vedena evidence stavu zásob. Celkový přehled je uveden v Kč a k dispozici jsou údaje z následujících dnů:

1. leden	20,523 mil. Kč
1. květen	16,100 mil. Kč
1. říjen	17,230 mil. Kč
1. leden	21,432 mil. Kč

Vypočtěte průměrný roční stav zásob (jejich hodnotu) v této firmě.

Ř e š e n í:

Protože vzdálenosti v měsících mezi časovými okamžiky jsou $d_1 = 4$, $d_2 = 5$, $d_3 = 3$, dostaneme

$$\bar{y}_{chr} = \frac{20,523\cdot 4 + 16,100(4+5) + 17,230(5+3) + 21,432\cdot 3}{2(4+5+3)} \doteq 17,880 \text{ mil. Kč.}$$

Průměrný roční stav zásob v podniku činil 17,880 mil. Kč. Tento průměrný roční stav by bylo možné také určit vzhledem k počtu dnů, resp. pracovních dnů.

Příklad 3.2

K dispozici jsou údaje o počtu zaměstnanců firmy METALIKA v průběhu kalendářního roku:

1. ledna	3 500 zaměstnanců
1. dubna	3 425 zaměstnanců
1. července	3 430 zaměstnanců
1. října	3 390 zaměstnanců
1. ledna	3 350 zaměstnanců

Vypočtěte průměrný počet zaměstnanců v dané firmě v celém ročním období.

Ř e š e n í:

Použijeme-li vztah pro výpočet chronologického průměru pro stejné časové vzdálenosti (v měsících), dostaneme

$$\bar{y}_{chr} = \frac{0,5.3\,500 + 3\,425 + 3\,430 + 3390 + 0,5.3\,350}{4} = 3\,417,5.$$

Chronologický průměr stavu zaměstnanců podniku METALIKA činí v daném roce 3 417,5.

Speciální typy řad

Součtové řady kumulativní

Kumulativní řady mají povahu narůstajících úhrnů a používají se u řad intervalového typu. Podstatou součtové řady kumulativní je načítání hodnot zvoleného jevu od stanoveného počátku. Tento nástroj má své uplatnění např. v případě sledování plnění ukazatelů za určité období (měsíc, rok). Kumulativní hodnoty nachází své uplatnění i v záležitostech strategického rozhodování. Ukázka použití tohoto nástroje je obsažena v následujícím příkladě.

Příklad 3.3
Máme údaje o stavu výroby (tis. tun) za jednotlivé měsíce. Úkolem je porovnání skutečného stavu výroby za jednotlivé měsíce a zároveň za celý rok:

Měsíc	Produkce (tis. tun)					
	Měsíční hodnoty			Kumulativní hodnoty		
	Plán	Výroba	(%)	Plán	Výroba	(%)
leden	36,2	36,5	100,8	36,2	36,5	100,8
únor	36,5	35,8	98,1	72,7	72,3	99,4
březen	35,1	35,8	102,0	107,8	108,1	100,3
duben	34,2	35,1	102,6	142,0	143,2	100,8
květen	33,0	33,2	100,6	175,0	176,4	100,8
červen	33,0	32,1	97,3	208,0	208,5	100,2
červenec	33,0	31,2	94,5	241,0	239,7	99,5
srpen	32,5	31,0	95,4	273,5	270,7	99,0
září	32,7	32,3	98,8	306,2	303,0	99,0
říjen	34,6	33,4	96,5	340,8	336,4	98,7
listopad	36,8	36,6	99,5	377,6	373,0	98,8
prosinec	38,0	38,1	100,3	415,6	411,1	98,9

Z tabulky kumulativních hodnot (kumulativních řad plánu a výroby) je vidět, jak se vyvíjí postupné plnění plánu (jde vždy o poměr dosažené výroby vzhledem k plánu v %) a z údajů za prosinec, vyplývá, že oproti ročnímu plánu 415,6 tis. tun bylo vyrobeno celkem za rok 411,1 tis. tun, což činí 98,9 % ročního plánu.

Časové řady kumulativních průměrů
Řady kumulativních průměrů jsou tvořeny z řad intervalových. Tyto řady ukazují, jak se kumulativní průměry blíží k celkovému průměru za sledované období, který je vyjádřen poslední hodnotou. Tento nástroj je využíván při sledování výše nákladů, při sledování kvality výroby atd. Princip použití vychází z kumulativní součtové řady, přičemž údaj se dělí počtem období, za které byl nakumulován. Pro demonstraci použijeme v následujícím příkladu data z příkladu 3.3 s kumulativními hodnotami výroby.

Příklad 3.4

Období	Kumulativní hodnota (tis. tun)	Kumulativní průměr (tis. tun)
leden	36,5	36,5 : 1 = 36,50
únor	72,3	72,3 : 2 = 36,15
březen	108,1	108,1 : 3 ≈ 36,03
duben	143,2	143,2 : 4 = 35,80
květen	176,4	176,4 : 5 = 35,28
červen	208,5	208,5 : 6 = 34,75
červenec	239,7	239,7 : 7 ≈ 34,24
srpen	270,7	270,7 : 8 ≈ 33,84
září	303,0	303,0 : 9 ≈ 33,67
říjen	336,4	336,4 :10 = 33,64
listopad	373,0	373,0 :11 ≈ 33,91
prosinec	411,1	411,1 :12 ≈ 34,26

Časové řady klouzavých součtů
Řady klouzavých součtů (úhrnů) mají charakter intervalových řad a získáme je postupnými součty předem zvoleného počtu po sobě jdoucích hodnot původní řady – viz příklad 3.5. Součtová řada klouzavých úhrnů je zejména vhodná ke srovnání vývojové tendence původní řady ve dvou delších (např. ročních) obdobích.

Časové řady klouzavých průměrů
Klouzavé průměry navazují na výpočet klouzavých součtů. Klouzavé průměry se vypočítají podělením klouzavého součtu počtem sečtených období - viz příklad 3.5. Řada klouzavých průměrů stírá případné sezónní vlivy na původní hodnoty, zejména extrémního charakteru. Další informace o klouzavých průměrech lze nalézt v [1], [3], [4] a [5].

Příklad 3.5
V následující tabulce jsou na časové řadě měsíčních údajů o výrobě z příkladu 3.3 demonstrovány klouzavé součty a klouzavé průměry pro 3 měsíce.

Měsíc	Výroba	Klouzavý součet	Klouzavý průměr
leden	36,5	---	---
únor	35,8	---	---
březen	35,8	108,1	36,033
duben	35,1	106,7	35,567
květen	33,2	104,1	34,700
červen	32,1	100,4	33,467
červenec	31,2	96,5	32,167
srpen	31,0	94,3	31,433
září	32,3	94,5	31,500
říjen	33,4	96,7	32,233
listopad	36,6	102,3	34,100
prosinec	38,1	108,1	36,033

Z obou vypočtených časových řad klouzavých součtů a klouzavých průměrů je zřetelně vidět, že od ledna do srpna měsíční výroba klesala a ve zbývajícím období rostla.

Vývoj časových řad

Mezi nejjednodušší charakteristiky rozboru časových řad patří absolutní a relativní míry růstu, respektive poklesu hodnot sledovaného znaku. Rozbor absolutních a relativních měr růstu umožňuje rozhodování pří výběru funkce na vyrovnání časové řady. U následujících nástrojů budeme pro jednoduchost uvažovat, že délky intervalů mezi sousedními okamžiky okamžikových časových řad, případně délky intervalů u intervalových časových řad, jsou stejné.

Absolutní míry růstu (růstu nebo poklesu) představují absolutní porovnání hodnot jednotlivých členů časové řady a pro bližší popis časové řady se používají:

absolutní přírůstek $\delta_i = y_i - y_{i-1}$ pro $i = 2,3,...,n$,

průměrný absolutní přírůstek $\bar{\delta} = \dfrac{1}{n-1}\sum_{i=2}^{n}\delta_i = \dfrac{y_n - y_1}{n-1}$,

který se vypočte jako prostý aritmetický průměr všech absolutních přírůstků.

Absolutní přírůstek se nazývá také ***první diference*** a značí se $\delta_i^{(1)}$. Pokud jsou absolutní přírůstky blízké konstantě, má hodnocená časová řada lineární trend, který lze graficky vyjádřit přímkou. Absolutní přírůstky charakterizují změny, tj. rychlost vývoje časové řady. Průměrný absolutní přírůstek pak vyjadřuje vývoj časové řady za celkové časové období. Zrychlení, resp. zpomalení vývoje časové řady popisují:

druhá diference $\delta_i^{(2)} = \delta_i - \delta_{i-1}$ pro $i = 3, 4, \ldots, n$,

průměrná druhá diference $\bar{\delta}^{(2)} = \dfrac{1}{n-2}\sum_{i=3}^{n}\delta_i^{(2)} = \dfrac{\delta_n - \delta_2}{n-2}$.

Jsou-li druhé diference blízké konstantě, je možné trend časové řady vyjádřit pomocí polynomu druhého stupně, tj. graficky parabolou. Obecně pokud je *m*–tá diference přibližně konstantní, lze průběh dané časové řady vyjádřit pomocí polynomu stupně *m*.

Poměrnou rychlost vývoje (růstu nebo poklesu) hodnot dané časové řady charakterizují relativní změny:

koeficient růstu $k_i = \dfrac{y_i}{y_{i-1}}$ pro $i = 2, 3, \ldots, n$,

koeficient přírůstku $\kappa_i = \dfrac{\delta_i}{y_{i-1}} = k_i - 1$ pro $i = 2, 3, \ldots, n$,

průměrný koeficient růstu $\bar{k} = \sqrt[n-1]{\prod_{i=2}^{n} k_i} = \sqrt[n-1]{\dfrac{y_n}{y_1}}$,

průměrný koeficient přírůstku $\bar{\kappa} = \bar{k} - 1$.

Koeficienty růstu a přírůstku i průměrné koeficienty růstu a přírůstku se také uvádějí v procentuálním tvaru, tj. krát 100. Pokud jsou koeficienty růstu přibližně konstantní, je průběh časové řady zhruba exponenciální.

Příklad 3.6

Vývoj hrubého domácího produktu (mld. Kč) v České republice v letech 1990 až 1996 je po přepočtu na stálé ceny v tabulce:

Rok	1990	1991	1992	1993	1994	1995	1996
HDP	564	721	859	1015	1192	1338	1579

Určete průměrný roční HDP, absolutní roční přírůstky, průměrný roční přírůstek, druhé diference, průměrnou druhou diferenci, koeficienty růstu, koeficienty přírůstku a průměrný koeficient růstu HDP.

Ř e š e n í:

Část výsledků výpočtu je v následující tabulce, kde místo *i* je přímo časová proměnná *t*:

t	y_t	δ_t	$\delta_t^{(2)}$	k_t	$k_t 100\%$	κ_t	$\kappa_t 100\%$
1990	564	---	---	---	---	---	---
1991	721	157	---	1,2784	127,84	0,2784	27,84
1992	859	138	-19	1,1914	119,14	0,1914	19,14
1993	1015	156	18	1,1816	118,16	0,1816	18,16
1994	1192	177	21	1,1744	117,44	0,1744	17,44
1995	1338	146	-31	1,1225	112,25	0,1225	12,25
1996	1579	241	95	1,1801	118,01	0,1801	18,01
Σ	7268	1015	84	---	---	---	---

Jedná se o intervalovou časovou řadu, takže průměrný roční HDP je

$$\bar{y} = \frac{7268}{7} \approx 1038,2857 \approx 1038,3 \text{ mld. Kč.}$$

Absolutní přírůstky, druhé diference, koeficienty růstu a koeficienty přírůstku jsou v předcházející tabulce. Odtud je zřejmé, že největšího HDP bylo dosaženo v roce 1996 a nejmenšího v roce 1990. Z tabulky dále vidíme, že největšího přírůstku HDP 241 mld. Kč bylo dosaženo v roce 1996 a naopak nejmenšího absolutního přírůstku HDP 138 mld. Kč bylo dosaženo v roce 1992. Avšak největšího relativního růstu HDP bylo dosaženo v roce 1991 (koeficient růstu je 1,2784, tedy 127,84 %) a nejmenšího relativního růstu HDP bylo dosaženo v roce 1995 (koeficient růstu je 1,1225, tedy 112,25 %). Z tabulky je dále vidět, že největšího absolutního zrychlení vývoje HDP (největší kladná druhá diference) bylo dosaženo v roce 1996 a největšího zpomalení vývoje HDP (největší záporná druhá diference) bylo dosaženo v roce 1995. Průměrný roční absolutní přírůstek HDP je

$$\bar{\delta} = \frac{1015}{7-1} = 169,1\bar{6} \doteq 169,2 \text{ mld. Kč,}$$

takže roční HDP celkově roste. Průměrný roční koeficient růstu HDP je

$$\bar{k} = \sqrt[7-1]{\frac{1579}{564}} \doteq \sqrt[6]{2,7996454} \doteq 1,1872,$$

tedy 118,72 %. Odtud je průměrný roční koeficient přírůstku HDP 0,1872, tedy 18,72 %. To opět odpovídá tomu, že roční HDP celkově roste. Poznamenejme, že výpočet průměrného koeficientu růstu nebo přírůstku pomocí aritmetického průměru je mírně řečeno zavádějící. To se ale bohužel v ekonomických aplikacích někdy stává. Průměrná roční druhá diference HDP je

$$\bar{\delta}^{(2)} = \frac{84}{7-2} = 16,8 \text{ mld. Kč,}$$

takže se růst HDP celkově zrychluje.

Popis časových řad

Při zkoumání vývoje sledovaného jevu v zákonitosti na čase nás kromě vývoje (růst, pokles, stagnace) zajímají zákonitosti časového vývoje. Vývoj časových řad je determinován kombinací několika vlivů působících na hodnoty časové řady. Jde o:

- ***trend vývoje*** (dlouhodobě působící vliv),
- ***periodické vlivy*** (pravidelně se opakující vliv),
- ***nahodilé vlivy*** (působí nepravidelně, resp. náhodně).

Trend časové řady

Trend je důležitý prvek časových řad a představuje obecnou tendenci dlouhodobého vývoje sledovaného ukazatele v čase. V rámci ekonomického využití časových řad je trend nejdůležitější složkou, která nás zajímá jak z hlediska současného stavu, tak i predikce budoucího vývoje. Často se u časové řady očekává lineární trend, vyjádřený lineární funkcí času *t* a graficky přímkou, ale v řadě případů jde o trend nelineárního tvaru. Pro vyjádření trendu časových řad byla vyvinuta a softwarově implementována řada metod [1], [3], [4], [5] a [6]. Základní metodika je níže popsána v odstavci o vyrovnání časových řad.

Periodické vlivy

Působením periodických vlivů dochází k periodickému kolísání průběhu časové řady. Délka periody je rozdílná a podle její velikosti uvažujeme další členění. Projevují se:

- ***cyklické vlivy*** (kolísání se opakuje pravidelně v jednotlivých letech dlouhého časového období),
- ***sezónní vlivy*** (kolísání se opakuje pravidelně v rámci jednoho delšího časového úseku - např.

měsíce v rámci roku),

- případně ***krátkodobé vlivy*** (kolísání krátkodobého charakteru při pravidelné periodě - např. den v týdnu, týden v měsíci atd.).

Pro vyjádření periodicity časových řad byla rovněž vyvinuta a softwarově implementována řada metod [1], [3], [4], [5] a [6].

Nahodilé vlivy

Nahodilé vlivy způsobují nahodilé výkyvy ukazatelů časových řad kolem trendu nebo trendu s periodickými výkyvy. Tyto vlivy považujeme za rušivou složku. Nahodilé vlivy jsou modelovány pomocí náhodných veličin a lze je diagnostikovat metodami matematické statistiky [1] a [2], z nichž část je popsána v kapitole 5.

Dekompozice časové řady

Z hlediska působení jednotlivých vlivů na průběh časové řady lze vyjádřit trend vývoje jako ***trendovou složku*** T_t, periodické vlivy jako ***periodickou složku*** P_t a náhodné vlivy jako ***náhodnou složku*** E_t dané časové řady. Periodickou složku podle potřeby rozdělujeme na ***cyklickou složku*** C_t a ***sezónní složku*** S_t. Dekompozice časové řady pak spočívá nejčastěji v jejím ***aditivním modelu***

$$y_t = T_t + P_t + E_t\,, \quad \text{resp. } y_t = T_t + C_t + S_t + E_t$$

anebo ***multiplikativním modelu***

$$\mathrm{y_t = T_t\, P_t\, E_t}\,, \quad \text{resp. } \mathrm{y_t = T_t\, C_t\, S_t\, E_t}\,.$$

Vyrovnání časových řad

Při zkoumání trendové složky časové řady jde vlastně o vymezení vlivu těch činitelů, které působí stabilně a určují směr vývoje dané časové řady. Graficky odpovídá řešení této úlohy nalezení takové dostatečně jednoduché křivky, která by při grafickém znázornění nejlépe vystihla směr vývoje dané časové řady. Takovou křivku získáme grafickým, mechanickým nebo analytickým vyrovnáním časové řady.

Grafické vyrovnání je založeno na zakreslení časové řady do grafu jako na obr. 3.1 a grafickým odhadem (vyrovnáním) jejího trendu ad hoc. Tato metoda má pouze orientační charakter, může vést k zavádějícím závěrům a díky software se již prakticky nepoužívá.

Mechanické vyrovnání časové řady vychází z klouzavých součtů. Když klouzavé součty dělíme počtem období, dostaneme klouzavé průměry, jejichž hodnoty jsou převážnou blízké původním hodnotám. Liší se tím, že jsou do určité míry zbavené sezónních výkyvů. Čára klouzavých průměrů bude tedy vyrovnanější než čára původních hodnot. Přitom je tím monotónnější (a zároveň kratší), čím více období vezmeme za základ pro stanovení příslušných klouzavých součtů. Velkou předností této metody je její jednoduchost a skutečnost, že nás dobře informuje o tendenci vývoje dané časové řady zbavené sezónních i cyklických výkyvů.

Analytické vyrovnání časové řady je založeno na předpokladu závislosti hodnot časové řady y_t na čase t. Pro vyrovnání časové řady používáme takovou funkci $f(t)$, která co nejlépe vyhovuje jejímu průběhu, tj. respektuje její trend, případně i její periodickou složku. Výběr vhodné funkce $f(t)$ je založen na rozboru průběhu původních empirických (pozorovaných) hodnot časové řady y_t, respektive jejich prvních, druhých, příp. dalších diferencí, a koeficientů růstu. Analytické vyrovnání časové řady má tvar

$$y_t = \hat{y}_t + e_t = f(t) + e_t\,,$$

kde $\hat{y}_t = f(t)$ je vyrovnaná hodnota pozorované hodnoty závisle proměnné y_t a e_t je tzv. ***reziduální složka*** v čase t. Pro analytické vyrovnání se v ekonomických aplikacích obvykle používají funkce, jejichž grafem je přímka, parabola, exponenciála, růstová křivka apod. Jde vlastně o aplikaci metod tzv. ***regresní analýzy*** [13], [17], [21] aj. V praxi se zejména u rozsáhlých časových řad provádí vyrovnání na PC pomocí statistického software.

Nejčastěji se používá při vyrovnání časové řady tzv. ***vyrovnání pomocí regresní přímky***, kdy předpokládáme, že řada má ***lineární trend***. Funkce $f(t)$ má tvar

$$\hat{y}_t = b_1 + b_2 t \,.$$

Koeficienty b_1 a b_2 stanovíme řešením tzv. ***soustavy normálních rovnic***

$$b_1 n + b_2 \sum_{t=1}^{n} t = \sum_{t=1}^{n} y_t \,,$$

$$b_1 \sum_{t=1}^{n} t + b_2 \sum_{t=1}^{n} t^2 = \sum_{t=1}^{n} y_t t \,.$$

První koeficient b_1 je y-ová souřadnice bodu, ve kterém daná přímka protíná osu y a odpovídá vyrovnané hodnotě časové řady v nultém období. Druhý koeficient b_2 je směrnice přímky a vyjadřuje samotný trend, to znamená sklon přímky. Odpovídá změně vyrovnaných hodnot $\hat{y}_t$ při jednotkové změně veličiny t a vyjadřuje průměrnou změnu původních hodnot y_t. O vhodnosti použité funkce se můžeme přesvědčit např. pomocí jejího grafu, velikosti koeficientu korelace nebo velikosti součtu čtverců odchylek (reziduí)

$$\sum_{t=1}^{n} (y_t - \hat{y}_t)^2 \,.$$

K výpočtu uvedené přímky můžeme také použít software Excel.

Příklad 3.7

Určete trendovou složku $\hat{y}_t = b_1 + b_2 t$ časové řady vývoje hrubého domácího produktu České republiky v letech 1990 až 1996 z příkladu 3.6.

Ř e š e n í:

Z příkladu 3.6 je n = 7, $\sum_{t=1}^{n} y_t = 7268$ a dalším výpočtem dostaneme $\sum_{t=1}^{n} t = 13591$, $\sum_{t=1}^{n} t^2 = 27804371$, $\sum_{t=1}^{n} y_t t = 14489736$. Soustava normálních rovnic pak je

$$7b_1 + 13591 b_2 = 7268 \,,$$

$$13591 b_1 + 27804371 b_2 = 14489736 \,.$$

Řešením této soustavy dostaneme koeficienty $b_1 \doteq -327237{,}3$ a $b_2 \doteq 164{,}7$. Odtud je vyrovnání časové řady (trendová funkce)

$$\hat{y}_t = -327237{,}3 + 164{,}7\, t \,.$$

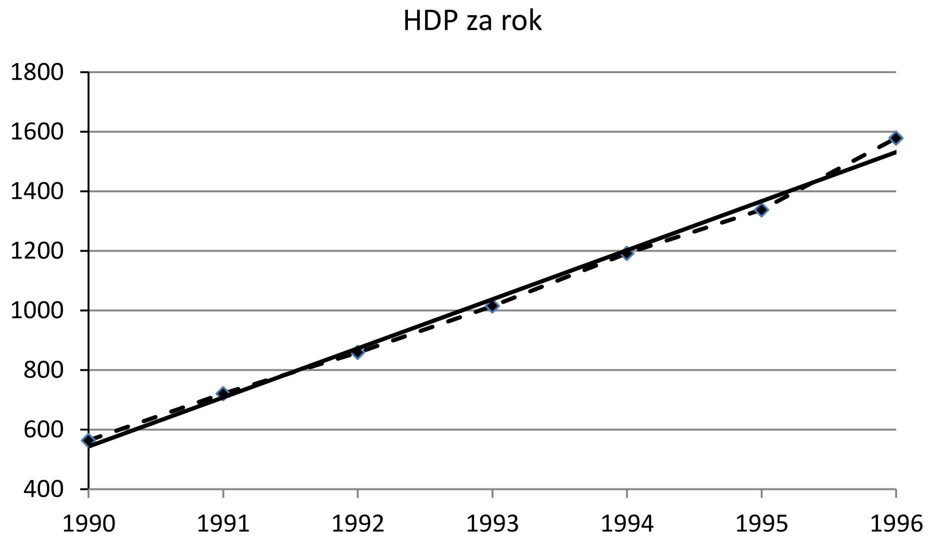
Například pro t = 1993 dostaneme

$$\hat{y}_{1993} = -327237{,}3 + 164{,}7 \cdot 1993 = 1009{,}8 \text{ mld. Kč},$$

což je v dobré shodě se skutečným HDP y_{1993} = 1015 mld. Kč. Také hodnota

b_2 = 164,7 mld. Kč / rok

odpovídá průměrnému ročnímu absolutnímu přírůstku $\overline{\delta} = 169{,}2$ mld. Kč / rok z příkladu 3.6. Koeficienty b_1 a b_2 můžeme také vypočítat na PC pomocí MS Excelu v nabídce Průvodce grafem nebo Analýza dat. Graf původní časové řady a trendové funkce je na obr. 3.3. Zřejmě je získaná trendová funkce vhodná.



Obr. 3.3

Kontrolní otázky

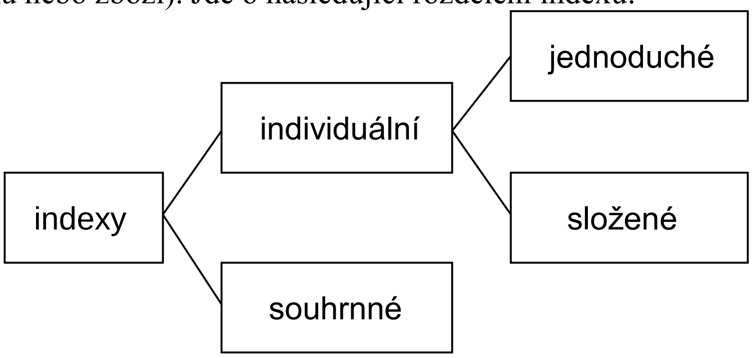
1. Definujte časovou řadu a uveďte konkrétní případy.
2. Uveďte rozdělení časových řad a konkrétní příklady na jednotlivé typy.
3. Jak se určí chronologický průměr okamžikové časové řady?
4. Jaké charakteristiky popisují vývoj časové řady?
5. Popište složky časové řady a její dekompozici.
6. Jakými způsoby vyrovnáváme časové řady?

4. INDEXOVÁ ANALÝZA

Základní pojmy

Indexy patří mezi poměrné kvantitativní statistické znaky a vyjadřují změnu sledovaného kvantitativního znaku nebo souboru znaků u jedné nebo více statistických jednotek během nějakého časového intervalu nebo vlivem nějakého faktoru. Konstruují se obvykle ve tvaru zlomku, kde v čitateli je hodnota znaku ve srovnávaném, tzv. ***běžném období***, a ve jmenovateli hodnota tohoto znaku v tzv. ***základním období***. Členění indexů vyplývá z charakteru, složitosti jejich konstrukce a vlivu zkoumaných veličin. Indexy dělíme na ***indexy individuální*** a ***indexy souhrnné***.

Individuální indexy vyjadřují změny velikosti jednotlivých znaků u jedné statistické jednotky anebo u skupiny ***stejnorodých*** statistických jednotek. Dělíme je na ***individuální jednoduché***, když popisujeme změnu znaku u jedné statistické jednotky (např. změnu ceny nebo množství jistého výrobku), a na ***individuální složené***, když sledujeme změnu jednoho znaku u skupiny podobných (homogenních) statistických jednotek (např. změnu cen nebo množství podobných výrobků). Oproti tomu ***souhrnné*** (***agregátní***) ***indexy*** vyjadřují změny velikosti znaků ***nestejnorodé*** skupiny typů statistických jednotek (např. ***spotřebního koše*** různých výrobků nebo zboží). Jde o následující rozdělení indexů:



Porovnávané znaky (veličiny) dělíme na ***extenzitní***, vyjadřující množství, objem, produkci apod., které značíme písmenem q a na ***intenzitní***, vyjadřující cenu, intenzitu apod., které značíme písmenem p. Z indexů pro extenzitní znaky se nejčastěji používají ***množstevní indexy*** a pro intenzitní znaky, povětšinou vyjadřující cenu, ***cenové indexy***. Indexy patří mezi nejčastěji aplikované popisné charakteristiky dynamiky vývoje makroekonomických i mikroekonomických statistických znaků.

Individuální indexy

Individuální jednoduché indexy
Individuální jednoduchý index i_q ***pro extenzitní znak***, resp. i_p ***pro intenzitní znak***, je

$$i_q = \frac{q_1}{q_0}, \text{ resp. } i_p = \frac{p_1}{p_0},$$

kde hodnota v čitateli odpovídá běžnému období a hodnota ve jmenovateli základnímu období. Velikost indexu závisí na velikosti čitatele a také na velikosti jmenovatele. Proto je důležitá správná volba základu. Za základ je třeba volit takové hodnoty, které nejlépe reprezentují daný soubor hodnot znaku. Někdy lze použít průměr z několika hodnot. Při výpočtu indexů a jejich následné interpretaci je třeba dbát srovnatelnosti období a také věcné srovnatelnosti ukazatelů. Při konstrukci indexu je třeba zajistit jednotu zkoumaných ukazatelů, jinak by vypovídací schopnost indexu byla značně snížena. V případě, že q vyjadřuje množství a p cenu, tak se ještě používá ***individuální jednoduchý hodnotový index***

$$i_h = \frac{q_1 p_1}{q_0 p_0} = i_q i_p \,.$$

Příklad 4.1

Produkce ocelárny byla v roce 1994 na úrovni 2780 tun, kdy cena za tunu oceli byla 8750,- Kč a v roce 1995 byla produkce oceli na úrovni 2950 tun, kdy cena za tunu oceli byla 9690,-Kč. Určete indexy množství, ceny a hodnoty produkce oceli v roce 1995 oproti roku 1994.

Ř e š e n í: $i_q = \frac{2950}{2780} \doteq 1{,}061 = 106{,}1\% \Rightarrow$ produkce ocelárny se meziročně zvýšila o 6,1%;

$i_p = \frac{9690}{8750} \doteq 1{,}107 = 110{,}7\% \Rightarrow$ cena oceli se meziročně zvýšila o 10,7%;

$i_h = \frac{q_1 p_1}{q_0 p_0} = i_q i_h \doteq 1{,}061 \cdot 1{,}107 \doteq 1{,}175 = 117{,}5\% \Rightarrow$ hodnota vyrobené oceli se meziročně zvýšila o 17,5%.

Individuální složené indexy

Pro posuzování změny sledovaného statistického znaku u skupiny podobných (homogenních) statistických jednotek používáme na rozdíl od jednoduchých individuálních indexů (např. index produkce cementu jedné cementárny) ***individuální složené indexy*** (např. index produkce skupiny cementáren, výstavu piva apod.). Rozlišení mezi jednoduchými a složenými indexy je velmi důležité. U složených indexů hraje roli srovnatelnost období, srovnatelnost věcná a srovnatelnost složení.

Individuální složený index extenzitní veličiny je dán vztahem

$$i_q = \frac{\sum_i q_1^{(i)}}{\sum_i q_0^{(i)}} \,.$$

Příklad 4.2

V následující tabulce je uvedena výroba cementu (v tunách) ve čtyřech cementárnách ve třech časových obdobích.

Cementárna	Leden (0)	Únor (1)	Březen (2)
A	2 300	2 450	2 390
B	5 210	5 100	4 800
C	8 100	8 600	9 000
D	6 250	6 250	5 900
Celková výroba	21 860	22 400	22 090

Jestliže považujeme všechny uvedené cementárny za jeden celek, pak individuální složené indexy pro extenzitní veličinu (produkci) jsou

$$i_{q1} = \frac{22\,400}{21\,860} \doteq 1{,}025 \;,\; i_{q2} = \frac{22\,090}{22\,400} \doteq 0{,}986 \;,$$

takže celková měsíční produkce cementáren se v únoru oproti lednu zvýšila o 2,5% a v březnu se oproti únoru snížila o 1,4%.

Individuální složený index charakterizující změnu intenzitní veličiny (např. cen) se nazývá ***index proměnlivého složení*** (***index průměrných cen***) a je dán vztahem

$$i_{prom.sloz.} = \frac{\overline{p}_1}{\overline{p}_0} = \frac{\dfrac{\sum_i p_1^{(i)} q_1^{(i)}}{\sum_i q_1^{(i)}}}{\dfrac{\sum_i p_0^{(i)} q_0^{(i)}}{\sum_i q_0^{(i)}}} \quad .$$

Individuální složený index vyjadřující změnu hodnoty celkové produkce, obratu, nákladů apod. je ***individuální složený hodnotový index***

$$I_h = \frac{\sum_i q_1^{(i)} p_1^{(i)}}{\sum_i q_0^{(i)} p_0^{(i)}} \, .$$

Příklad 4.3

V následující tabulce je uveden přehled vývoje cen a množství dodaných (vyrobených) výrobků stejného druhu v podniku. Poslední dva vypočtené sloupce obsahují hodnoty dodávek.

Dodávka	Počet kusů v období		Cena za kus v období		Hodnota dodávky v období	
	Základní	Běžné	Základní	Běžné	Základní	Běžné
	q_0	q_1	p_0	p_1	p_0q_0	p_1q_1
Smluvní	8 000	8 000	12,-	13,2	96 000	105 600
Nadsmluvní	500	4 600	30,-	30,-	15 000	138 000
Celkem	8 500	12 600	-----	-----	111 000	243 600

Intenzitní veličinou, jejíž změnu zkoumáme, je cena. Změnu ceny pro jednotlivé druhy dodávek charakterizují individuální jednoduché cenové indexy pro smluvní dodávky (1) a pro nadsmluvní dodávky (2):

$$i_p^{(1)} = \frac{p_1^{(1)}}{p_0^{(1)}} = \frac{13,2}{12} = 1{,}100 \ , \qquad i_p^{(2)} = \frac{p_1^{(2)}}{p_0^{(2)}} = \frac{30}{30} = 1{,}000 \, .$$

U smluvních dodávek stoupla cena o 10% a u nadsmluvních zůstala stejná. Extenzitní veličinou, jejíž změnu zkoumáme, je velikost dodávek. Změnu množství v jednotlivých druzích dodávek charakterizují individuální jednoduché množstevní indexy pro smluvní dodávky (1) a pro nadsmluvní dodávky (2):

$$i_q^{(1)} = \frac{q_1^{(1)}}{q_0^{(1)}} = \frac{8000}{8000} = 1{,}000 , \qquad i_q^{(2)} = \frac{q_1^{(2)}}{q_0^{(2)}} = \frac{4600}{500} = 9{,}200 .$$

U smluvních dodávek zůstalo množství stejné a u nadsmluvních se zvýšilo o 820%.

Změny celkových dodávek vyjadřuje individuální složený množstevní index

$$i_q = \frac{\sum_i q_1^{(i)}}{\sum_i q_0^{(i)}} = \frac{12600}{8500} = 1{,}482 = 148{,}2\% \ ,$$

takže celkové množství dodaných výrobků vzrostlo o 48,2%.

Průměrná cena v základním období je

$$\bar{p}_0 = \frac{\sum p_0^{(i)} q_0^{(i)}}{\sum q_0^{(i)}} = \frac{111\,000}{8\,500} \doteq 13,06 \text{ Kč}$$

a průměrná cena v běžném období je

$$\bar{p}_1 = \frac{\sum p_1^{(i)} q_1^{(i)}}{\sum q_1^{(i)}} = \frac{243\,600}{12\,600} \doteq 19,33 \text{ Kč},$$

takže index proměnlivého složení (průměrných cen) je

$$i_{prom.sloz.} = \frac{\bar{p}_1}{\bar{p}_0} \doteq \frac{19,33}{13,06} \doteq 1,480 = 148,0\%,$$

což znamená, že průměrné ceny dodaných výrobků vzrostly o 48,0%.

Individuální jednoduché hodnotové indexy jsou

$$i_h^{(1)} = i_q^{(1)} i_p^{(1)} = 1,000 \cdot 1,100 = 1,100 = 110,0\%,$$

$$i_h^{(2)} = i_q^{(2)} i_p^{(2)} = 9,200 \cdot 1,000 = 9,200 = 920,0\%$$

a individuální složený hodnotový index dodávek je

$$I_h = \frac{\sum_i q_1^{(i)} p_1^{(i)}}{\sum_i q_0^{(i)} p_0^{(i)}} = \frac{243600}{111000} \doteq 2,195,$$

takže celková hodnota dodávek se zvětšila o 119,5%.

Souhrnné indexy

Základní vlastností ***souhrnných*** (***agregátních***) ***indexů*** je, že s jejich pomocí lze měřit změny jak extenzitních, tak i intenzitních veličin nesourodých (heterogenních) statistických znaků a veličin. Jestliže se změní např. ceny spotřebních produktů, zajímá nás, o kolik procent poklesly ceny u produktů jako celku. Protože jde o různé druhy zboží, nelze počítat průměrnou cenu připadající na jednotný výrobek.

Kombinovaný efekt intenzitních veličin a extenzitních veličin na změnu celkové hodnoty produktu vyjadřuje tzv. ***souhrnný hodnotový index*** (**index obratu**, ***index nákladů***)

$$I_h = \frac{\sum_i q_1^{(i)} p_1^{(i)}}{\sum_i q_0^{(i)} p_0^{(i)}},$$

kde extenzitními znaky jsou např. množství, počty apod. a intenzitními znaky jsou např. ceny, jednotkové náklady apod.

Abychom zjistili vliv jen jednoho činitele, je třeba vliv druhého eliminovat, což provedeme jeho ustálením na zvolené úrovni v obou porovnávaných agregátech. Stálou úroveň intenzitní (p) nebo extenzitní (q) veličiny v indexu dosahujeme obvykle ustálením na úrovni základního anebo ustálením na úrovni běžného období. Podle toho dostaneme:

- ***Laspeyresův souhrnný index pro extenzitní veličinu*** $I_q^L = \frac{\sum_i q_1^{(i)} p_0^{(i)}}{\sum_i q_0^{(i)} p_0^{(i)}}$,
- ***Laspeyresův souhrnný index pro intenzitní veličinu*** $I_p^L = \frac{\sum_i q_0^{(i)} p_1^{(i)}}{\sum_i q_0^{(i)} p_0^{(i)}}$,

- ***Paascheho souhrnný index pro extenzitní veličinu*** $I_q^P = \dfrac{\sum_i q_1^{(i)} p_1^{(i)}}{\sum_i q_0^{(i)} p_1^{(i)}}$,

- ***Paascheho souhrnný index pro intenzitní veličinu*** $I_p^P = \dfrac{\sum_i q_1^{(i)} p_1^{(i)}}{\sum_i q_1^{(i)} p_0^{(i)}}$.

Mezi uvedenými souhrnnými indexy extenzitní veličiny a intenzitní veličiny a souhrnným hodnotovým indexem platí vztah

$$I_h = I_p^L\, I_q^P = I_q^L\, I_p^P .$$

Laspeyresův typ indexu ani Paascheho typ indexu nevyjadřují vždy dostatečně změnu, neboť změna veličin p, resp. q mezi obdobím základním a obdobím běžným ovlivňuje změnu veličin q, resp. p. Např. změna cen ovlivňuje spotřebu a naopak. Pro eliminaci tohoto nedostatku se nejčastěji užívá tzv. ***Fisherův ideální index***

$$I_q^F = \sqrt{I_q^L\, I_q^P}\text{ , resp. } I_p^F = \sqrt{I_p^L\, I_p^P} .$$

Vzhledem k tomu, že ani Fisherův index nepopisuje změny dostatečně výstižně (je pouze kompromisem mezi Laspeyresovým a Paascheho indexem, jejich geometrickým průměrem), používají se další indexy [1].

Příklad 4.4

Jsou dány údaje o prodeji nestejnorodých výrobků A, B, C obchodní společnosti:

Produkt	Prodané množství		Maloobchodní cena	
	Základní období	Běžné období	Základní období	Běžné období
	q_0	q_1	p_0	p_1
A	1 000	1 200	60	69
B	6 000	4 500	10	11
C	8 000	9 000	8	7

Stanovte: a) růst celkových tržeb společnosti v daném období,
b) růst objemu prodeje,
c) změnu ceny prodaného zboží.

Ř e š e n í:

V následující tabulce jsou uvedeny pomocné výpočty:

Produkt	q_0p_0	q_1p_1	q_0p_1	q_1p_0
A	60 000	82 800	69 000	72 000
B	60 000	49 500	66 000	45 000
C	64 000	63 000	56 000	72 000
Celkem	184 000	195 300	191 000	189 000

Z vypočtených hodnot uvedených v tabulce je hodnotový index

$$I_h = \frac{\sum_i q_1^{(i)} p_1^{(i)}}{\sum_i q_0^{(i)} p_0^{(i)}} = \frac{195\,300}{184\,000} \doteq 1{,}0614 = 106{,}14\%,$$

takže růst celkových tržeb společnosti je 6,14%.

Laspeyresův index objemu prodaného zboží je

$$I_q^L = \frac{\sum_i q_1^{(i)} p_0^{(i)}}{\sum_i q_0^{(i)} p_0^{(i)}} = \frac{189\,000}{184\,000} \doteq 1{,}0272 = 102{,}72\%$$

a Laspeyresův index cen prodaného zboží je

$$I_p^L = \frac{\sum_i q_0^{(i)} p_1^{(i)}}{\sum_i q_0^{(i)} p_0^{(i)}} = \frac{191\,000}{184\,000} \doteq 1{,}0380 = 103{,}80\%,$$

takže podle Laspeyersova indexu vzrostl objem prodaného zboží o 2,72% a cena prodaného zboží vzrostla o 3,80%.

Paascheho index objemu prodaného zboží je

$$I_q^P = \frac{\sum_i q_1^{(i)} p_1^{(i)}}{\sum_i q_0^{(i)} p_1^{(i)}} = \frac{195\,300}{191\,000} \doteq 1{,}0225 = 102{,}25\%,$$

a Paascheho index cen prodaného zboží je

$$I_p^P = \frac{\sum_i q_1^{(i)} p_1^{(i)}}{\sum_i q_1^{(i)} p_0^{(i)}} = \frac{195\,300}{189\,000} \doteq 1{,}0333 = 103{,}33\%,$$

takže podle Paascheho indexu vzrostl objem prodaného zboží o 2,25% a cena prodaného zboží vzrostla o 3,33%.

Fisherův index objemu prodaného zboží je

$$I_q^F = \sqrt{I_q^L\, I_q^P} \doteq \sqrt{1{,}0272 \cdot 1{,}0225} \doteq 1{,}0248 = 102{,}48\%,$$

a Fisherův index cen prodaného zboží je

$$I_p^F = \sqrt{I_p^L\, I_p^P} \doteq \sqrt{1{,}0380 \cdot 1{,}0333} \doteq 1{,}0356 = 103{,}56\%,$$

takže podle Fisherova indexu vzrostl objem prodaného zboží o 2,25% a cena prodaného zboží vzrostla o 3,33%.

Poznamenejme ještě, že pro každý druh zboží A, B a C bychom mohli také vypočítat individuální jednoduchý množstevní, cenový a hodnotový index, tj. celkem dalších 9 indexů.

Indexy a absolutní veličiny

Indexy vyjadřují jenom poměrné změny zkoumaných jevů. Proto samy o sobě nemohou stačit k rozboru vývoje těchto jevů. K tomuto účelu je třeba znát i absolutní změnu veličiny, z nichž byl index vypočítán. U pozorovaných veličin a jejich agregátů nás vedle indexů zajímají také jejich absolutní přírůstky, které vyjadřují rozdíly čitatelů a jmenovatelů ve vzorcích pro indexy, tj. $q_1 - q_0$, $p_1 - p_0$ apod.

Bázické a řetězové indexy

Při zkoumání časového vývoje reálných jevů je často vhodné vytvořit celou řadu (posloupnost) indexů za několik po sobě jdoucích období. Používají se zejména následující dva typy indexů, kde y_n značí extenzitní nebo intenzitní veličiny, příp. jejich agregáty.

Jedno období je považováno za ***základní*** s hodnotou sledované veličiny y_0 a s ní jsou porovnány hodnoty sledované veličiny y_n v ***běžných*** (ostatních) obdobích. Za základní období můžeme přitom zvolit libovolné (ale jediné) ze sledovaných období. Získáme tak ***indexy bázické*** neboli ***indexy se stálým základem***

$$i_{n/0} = \frac{y_n}{y_0}, \quad n = 1, 2, \ldots .$$

Dané období porovnáváme s předcházejícím obdobím, tedy hodnotu sledované veličiny y_n s hodnotou y_{n-1}. Získáme tak ***indexy řetězové*** neboli ***indexy s pohyblivým základem***

$$i_{n/n-1} = \frac{y_n}{y_{n-1}}, \quad n = 1, 2, \ldots .$$

Řetězové indexy jsou speciálním případem tzv. ***meziobdobních indexů***

$$i_{n/n-k} = \frac{y_n}{y_{n-k}},$$

kde za číslo k volíme nejčastěji délku periody (cyklu nebo sezóny).

Z bázických i řetězových indexů často vytváříme časové řady, které zpracováváme metodami popsanými v kapitole 3. Např. geometrický průměr z řetězových indexů představuje průměrný index vyjadřující stejnou poměrnou změnu sledovaného znaku mezi jednotlivými stejně dlouhými obdobími. Jestliže bychom použili místo geometrického průměru aritmetický průměr, dopustili bychom se zásadní chyby.

Příklad 4.5

Vypočtěte bázické a řetězové indexy a měsíční přírůstky pro údaje v následující tabulce:

Období	Produkce	Označení
leden	75 000 t	q_0
únor	75 250 t	q_1
březen	81 000 t	q_2
duben	82 100 t	q_3

Ř e š e n í:

Bázické indexy: $i_{1/0} \approx 1{,}0033$; $i_{2/0} = 1{,}0800$; $i_{3/0} \approx 1{,}0947$

Řetězové indexy: $i_{1/0} \approx 1{,}0033$; $i_{2/1} \approx 1{,}0764$; $i_{3/2} \approx 1{,}0136$

Měsíční přírůstky: 250 t; 5 750 t; 1100 t

Kontrolní otázky

1. Definujte pojem indexu a popište jeho typy.
2. Jak se určují individuální jednoduché indexy a jaké mají vlastnosti?
3. Jak se určují individuální složené indexy a jaké mají vlastnosti?
4. Jak se určují souhrnné indexy a jaké mají vlastnosti?
5. Definujte bázické a řetězové indexy a popište jejich použití.
6. Specifikujte nedostatky Laspeyresova a Paascheho indexu a ilustrujte je na konkrétním spotřebním koši.

5. LITERATURA

Učebnice a monografie

1. Aczel, A. D. *Complete Business Statistics*. Chicago : IRWIN, 1989.
2. Anděl, J. *Základy matematické statistiky.* Praha: MATFZYPRESS, 2005.
3. Bowerman, B. L. - O´Connell, R. T. *Applied Statistics - Improving Business Processes*. Chicago : IRWIN, 1997.
4. Cyhelský, L. - Kahounová, J. - Hindls, R. *Elementární statistická analýza.* 1. vyd. Praha : Management Press, 1996.
5. Hahn, G. J. - Shapiro, S. S. *Statistical Models in Engineering*. New York : John Wiley & Sons, Inc., 1994.
6. Budíková, M., Králová, M., Maroš, B. *Průvodce základními statistickými metodami.* Praha: Grada Publishing, 2010.
7. Kožíšek, J., Stieberová, B. *Statistika v příkladech*. Praha: Verlag Dashöfer, 2012.
8. Lamoš, F. - Potocký, R. *Pravdepodobnosť a matematická štatistika*. 1. vyd. Bratislava : ALFA, 1989.
9. Likeš, J. - Machek, J. *Matematická statistika*. 1. vyd. Praha : SNTL, 1983.
10. Meloun, M. a Militký, J. *Statistické zpracování experimentálních dat.* Praha : PLUS, 1994.
11. Montgomery, D. C. - Renger, G. *Probability and Statistics*. New York : John Wiley & Sons, Inc., 2003.
12. Potocký, R. et. al. *Zbierka úloh z pravdepodobnosti a matematickej štatistiky.* 1. vyd. Bratislava : ALFA/SNTL, 1986.
13. Seger, J. - Hindls, R. *Statistické metody v tržním hospodářství*. 1. vyd. Praha : Victoria Publishing, 1995.
14. Swoboda, H. *Moderní statistika*. 1. vyd. Praha : Svoboda, 1977.
15. Sprinthall, R. C. *Basic Statistical Analysis*. 5th ed. Boston : Allyn and Bacon, 1997.
16. Triola, M. F. *Elementary Statistics*. Redwood City : B/C Publishing Comp., 1989.
17. Wonnacot, T. H. - Wonnacot, R. J. *Statistika pro obchod a hospodářství*. 1. vyd. Praha : Victoria Publishing, 1993.
18. Zvára, K. - Štěpán, J. *Pravděpodobnost a matematická statistika*. 1. vyd. Praha : Matfyzpress, 1997.

Učební texty

19. Budíková, M. - Mikoláš, Š. - Osecký, P. *Teorie pravděpodobnosti a matematická statistika - Sbírka příkladů.* 1. vyd. Brno : MU, 1996.
20. Jarošová, E. *Statistika B - Řešené příklady*. 1. vyd. Praha : VŠE, 1994.
21. Karpíšek, Z. *Matematika IV - Statistika a pravděpodobnost.* 4 vyd. Brno : FSI VUT v Akademickém nakladatelství CERM, s. r. o. Brno, 2003.
22. Karpíšek, Z. - Drdla, M. *Applied Statistics*. 1. vyd. Brno : FP VUT v PC - DIR, 1999.
23. Karpíšek, Z. – Popela, P. – Bednář, J. *Statistika a pravděpodobnost. Učební pomůcka - studijní opora pro kombinované studium.* Brno : FSI VUT v CERM Brno, 2002.
24. Koutková, H. - Moll, I. *Úvod do pravděpodobnosti a matematické statistiky.* 1. vyd. Brno : ES VUT, 1990.
25. Kropáč, J. *Úvod do počtu pravděpodobnosti a matematické statistiky.* 1. vyd. Brno : VA, 2000.
26. Likeš, J. - Cyhelský, L. - Hindls, R. *Úvod do statistiky a pravděpodobnosti - Statistika A.* 1. vyd. Praha : VŠE, 1995.
27. Reif, J. *Metody matematické statistiky*. 1. vyd. Plzeň : Západočeská univerzita, 2000.
28. Sadovský, Z. *Statistická analýza I,* 1. vyd. Brno, VUT, 1982
29. Sadovský, Z. *Statistická analýza II,* 1. vyd. Brno, VUT, 1986
30. Seberová, H. *Statistika I, II*. 1. vyd. Vyškov : VVŠ PV, 1995.

WWW odkazy

31. http://badame.vse.cz/
32. http://iastat.vse.cz/
33. http://davidmlane.com/hyperstat/
34. http://home.zcu.cz/~friesl/Vyuka/Odkazy.html
35. http://www.math.csusb.edu/faculty/stanton/m262/index.html
36. http://onlinestatbook.com/rvls.html
37. http://www.stat.sc.edu/rsrch/gasp/
38. http://www.statsoft.com/textbook/stathome.html
39. http://www.statsoft.cz/
40. http://www.trilobyte.cz/

DODATEK – ELEMENTY TEORIE PRAVDĚPODOBNOSTI

Náhodné jevy

Náhodný jev je výsledek ***pokusu*** (tj. realizace určitého systému podmínek) a jeho charakteristickým rysem je, že může, ale nemusí nastat. Míru možnosti jeho nastoupení vyjadřuje v číselné formě jeho ***pravděpodobnost***. U náhodných jevů požadujeme ***hromadnost*** a ***stabilitu***, tj. dostatečnou opakovatelnost a neměnnost pokusu. Nezbytným předpokladem je také ***rozpoznatelnost*** náhodných jevů.

Jednotlivým možným (uvažovaným) výsledkům pokusu odpovídají ***elementární náhodné jevy***, které vyjadřujeme pomocí jednoprvkových množin označených symbolem $\{\omega\}$. Všechny možné výsledky pokusu pak tvoří množinu Ω (sestávající z prvků ω), kterou nazýváme ***základní prostor***, takže $\omega \in \Omega$. ***Náhodným jevem*** A pak rozumíme libovolnou podmnožinu základního prostoru Ω, tedy $A \subseteq \Omega$. Při pokusu nastane vždy jeden libovolný elementární náhodný jev $\{\omega\}$ a s ním současně nastanou všechny náhodné jevy, které jej obsahují, a naopak nenastanou všechny náhodné jevy, které jej neobsahují. Speciálním případem je ***jistý jev***, který nastane při každém pokusu, a ***nemožný jev***, který nenastane při žádném pokusu. Jistý jev je proto ekvivalentní základnímu prostoru Ω a nemožný jev naopak prázdné množině $\varnothing$.

Vztahy mezi náhodnými jevy vyjadřujeme pomocí množinových inkluzí:

a) $A \subseteq B$ znamená, že nastoupení náhodného jevu A ***má za následek*** nastoupení náhodného jevu B.

b) $A = B$ značí ***rovnost*** (***ekvivalenci***) náhodných jevů A a B.

Operace s náhodnými jevy vyjadřujeme pomocí množinových operací:

a) ***Průnik*** náhodných jevů $A \cap B$ nastane, jestliže nastanou oba náhodné jevy A a B. Analogicky definujeme náhodné jevy $\bigcap_{i=1}^{n} A_i$ a $\bigcap_{i=1}^{\infty} A_i$, které nastanou, jestliže nastanou všechny náhodné jevy A_i .

b) ***Sjednocení*** náhodných jevů $A \cup B$ nastane, jestliže nastane aspoň jeden z náhodných jevů A a B, tedy A nebo B. Analogicky definujeme náhodné jevy $\bigcup_{i=1}^{n} A_i$ a $\bigcup_{i=1}^{\infty} A_i$,které nastanou, jestliže nastane aspoň jeden náhodný jev A_i.

c) ***Rozdíl*** náhodných jevů $A - B$ nastane, jestliže nastane náhodný jev A a nenastane náhodný jev B.

d) ***Opačný*** náhodný jev k náhodnému jevu A je jev $\bar{A} = \Omega - A$, který nastane, jestliže nenastane náhodný jev A.

e) Náhodné jevy A a B jsou ***disjunktní***, jestliže $A \cap B = \varnothing$. Náhodné jevy A_i, $i = 1,2, \dots$ jsou ***disjunktní***, jestliže jsou disjunktní všechny dvojice náhodných jevů A_i, A_j pro $i \neq j$.

Vlastnosti operací s náhodnými jevy jsou samozřejmě totožné s vlastnostmi operací s množinami. Abychom mohli definovat pravděpodobnost náhodného jevu, zabýváme se jenom takovými náhodnými jevy na Ω, které tvoří následující strukturu.

Jevové pole Σ na Ω je množina náhodných jevů (systém podmnožin základního prostoru Ω) s vlastnostmi:

1. $\varnothing \in \Sigma$, $\Omega \in \Sigma$.
2. Pro každý náhodný jev $A \in \Sigma$ je $\bar{A} \in \Sigma$.

3. Pro každou posloupnost náhodných jevů $A_i \in \Sigma$, $i = 1, 2, \ldots$, je $\bigcap_{i=1}^{\infty} A_i \in \Sigma$.

Příklad 1
Náhodný pokus spočívá v jednom hodu šestistěnnou hrací kostkou se stěnami očíslovanými od 1 do 6. Náhodný jev A nastoupí, jestliže padne sudé číslo a náhodný jev B nastoupí, jestliže padne číslo větší než 4. Určete Ω, $\overline{A}$, $\overline{B}$, $A \cup B$, $A \cap B$, $A - B$, $B - A$, Σ.
Ř e š e n í:
Základní prostor je $\Omega = \{1, 2, 3, 4, 5, 6\}$ je konečný a elementární náhodné jevy jsou $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$. Dále je $A = \{2, 4, 6\}$ a $B = \{5, 6\}$, takže

$\overline{A} = \{1, 3, 5\}$ ... padne liché číslo,
$\overline{B} = \{1, 2, 3, 4\}$ ... padne číslo menší než 5,
$A \cup B = \{2, 4, 6\} \cup \{5, 6\} = \{2, 4, 5, 6\}$ ... nepadne číslo 1 a 3,
$A \cap B = \{2, 4, 6\} \cap \{5, 6\} = \{6\}$ ... padne číslo 6,
$A - B = \{2, 4, 6\} - \{5, 6\} = \{2, 4\}$ ... padne číslo 2 nebo 4,
$B - A = \{5, 6\} - \{2, 4, 6\} = \{5\}$ ... padne číslo 5.

Protože nejsou stanovena žádná omezení na náhodné jevy, můžeme uvažovat maximální jevové pole (množinu všech podmnožin konečného základního prostoru Ω)

$$\Sigma = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \ldots, \{5, 6\}, \ldots, \{2, 3, 4, 5, 6\}, \Omega\},$$

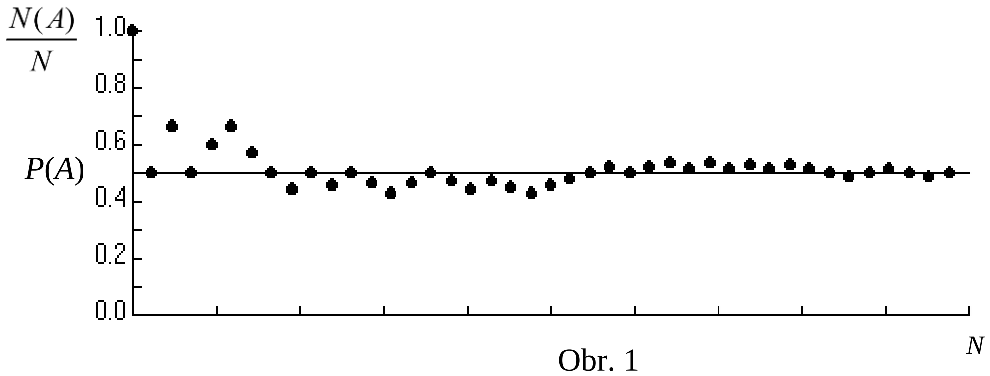
které obsahuje $2^6 = 64$ náhodných jevů.

Pravděpodobnost a její vlastnosti

Jestliže při opakovaných sériích náhodných pokusů, které sestávají vždy z N pokusů, sledujeme chování ***relativní četnosti*** nastoupení náhodného jevu A, tj. posloupností čísel

$$\frac{N(A)}{N},$$

kde $N(A)$ je počet nastoupení jevu A v dané sérii N pokusů, pak vidíme, že posloupnosti relativních četností mají ve skoro všech sériích snahu konvergovat pro dostatečně velký počet pokusů N k jisté pevné hodnotě $P(A)$ – viz příklad jedné takové posloupnosti na obr. 1.



Obr. 1

Tato teoretická hodnota $P(A)$ vyjadřuje míru možnosti nastoupení náhodného jevu A v jednotlivém pokusu a hovoříme o tzv. „***statistické definici pravděpodobnosti***“ náhodného jevu A.

Z jakékoliv realizované série N pokusů však můžeme pravděpodobnost $P(A)$ náhodného jevu A pomocí zjištěné relativní četnosti $\frac{N(A)}{N}$ pouze více či méně přesně odhadnout. Naopak pravděpodobnost $P(A)$ znamená, že při mnoha pokusech (řádově tisíce a více) nastoupí náhodný jev A zhruba ve $100P(A)$ % pokusů. Na vlastnostech relativní četnosti

$$0 \leq \frac{N(A)}{N} \leq 1,$$

$$A \cap B = \varnothing \;\Rightarrow\; \frac{N(A \cup B)}{N} = \frac{N(A)}{N} + \frac{N(B)}{N},$$

je založena následující obecná (axiomatická) definice pravděpodobnosti náhodného jevu.

Pravděpodobnost $P(A)$ náhodného jevu $A \in \Sigma$ je reálná funkce definovaná na jevovém poli Σ s vlastnostmi:

1. $P(A) \geq 0$ pro všechny náhodné jevy $A \in \Sigma$.
2. $P(\Omega) = 1$.
3. Pro každou posloupnost disjunktních náhodných jevů $A_i \in \Sigma$, $i = 1, 2, \ldots$, je

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Uspořádaná trojice (Ω, Σ, P) se nazývá ***pravděpodobnostní prostor***.

Pravděpodobnost má tyto základní vlastnosti:

a) $P(\overline{A}) = 1 - P(A)$, $P(\varnothing) = 0$, $0 \leq P(A) \leq 1$,

b) $A \subseteq B \Rightarrow P(A) \leq P(B)$, $P(B - A) = P(B) - P(A)$,

c) $P(A_1 \cup \ldots \cup A_n) = 1 - P(\overline{A}_1 \cap \ldots \cap \overline{A}_n) =$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i,j=1 \\ i<j}}^{n} P(A_i \cap A_j) + \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n), \quad n \geq 2;$$

speciálně pro $n = 2$ je

$P(A \cup B) = 1 - P(\overline{A} \cap \overline{B}) = P(A) + P(B) - P(A \cap B)$.

d) Jestliže základní prostor Ω je konečný nebo spočetný (tj. elementární jevy $\{\omega\}$ lze uspořádat do posloupnosti), pak

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

a pro základní prostor Ω tvořený n stejně pravděpodobnými elementárními jevy $\{\omega\}$ je

$$P(A) = \frac{m}{n},$$

kde m je počet elementárních jevů $\{\omega\}$, z nichž sestává náhodný jev A. Říkáme, že m je ***počet příznivých výsledků pokusu*** a n je ***počet možných výsledků pokusu***. Hovoříme přitom o tzv. „***klasické definici pravděpodobnosti***“ náhodného jevu A.

Příklad 2

Vypočtěte pravděpodobnosti $P(A)$, $P(B)$, $P(\overline{A})$, $P(\overline{B})$, $P(A \cup B)$, $P(A \cap B)$,
$P(A - B)$, $P(B - A)$ náhodných jevů z příkladu 1 pro hrací kostku
z homogenního materiálu ve tvaru krychle.

Ř e š e n í:
Všechny elementární náhodné jevy mají vzhledem k pravidelnosti a homogennosti hrací kostky stejnou pravděpodobnost $P(\{\omega\}) = \frac{1}{6}$ a $n = 6$. Přímým výpočtem z „klasické definice pravděpodobnosti“ obdržíme

$$P(A) = \frac{3}{6} = \frac{1}{2}, \qquad P(B) = \frac{2}{6} = \frac{1}{3}, \qquad P(\bar{A}) = \frac{3}{6} = \frac{1}{2}, \qquad P(\bar{B}) = \frac{4}{6} = \frac{2}{3},$$

$$P(A \cup B) = \frac{4}{6} = \frac{2}{3}, \quad P(A \cap B) = \frac{1}{6}, \qquad P(A - B) = \frac{2}{6} = \frac{1}{3}, \quad P(A - B) = \frac{1}{6}.$$

Z vlastností pravděpodobnosti lze např. určit

$$P(\bar{A}) = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2},$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}.$$

Příklad 3
Víme, že v dodávce 100 hřídelí nemá požadovaný průměr 10 kusů, požadovanou délku nemá 20 kusů a současně nemá požadovaný průměr i délku 5 kusů. Určete pravděpodobnost toho, že náhodně vybraná hřídel z dodávky má požadovaný průměr i délku.
Ř e š e n í:
Předpokládejme, že každá hřídel v dodávce má stejnou pravděpodobnost výběru. Jestliže A značí náhodný jev, že vybraná hřídel nemá požadovaný průměr a B značí náhodný jev, že vybraná hřídel nemá požadovanou délku, pak

$$P(A) = 10/100 = 0{,}10, \; P(B) = 20/100 = 0{,}20,$$

$$P(A \cap B) = 5/100 = 0{,}05.$$

Pravděpodobnost toho, že náhodně vybraná hřídel má požadovaný průměr i délku, je

$$P(\bar{A} \cap \bar{B}) = 1 - P(\overline{\bar{A} \cap \bar{B}}) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)] =$$
$$= 1 - (0{,}10 + 0{,}20 - 0{,}05) = 0{,}75 \; .$$

Podmíněná pravděpodobnost a nezávislé jevy

Pravděpodobnost náhodného jevu $A \in \Sigma$ za podmínky (předpokladu), že nastane náhodný jev $B \in \Sigma$, $P(B) \neq 0$, je ***podmíněná pravděpodobnost***

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \; .$$

Podmíněná pravděpodobnost $P(A|B)$ je relativní mírou nastoupení náhodného jevu A vzhledem k míře možnosti nastoupení náhodného jevu B. Podmíněná pravděpodobnost má základní vlastnosti:

a) $P(A_1 \cap \ldots \cap A_n) = P(A_1)(A_2 \mid A_1) \cdots P(A_n \mid A_1 \cap \ldots \cap A_{n-1})$,
speciálně pro $n = 2$ je $P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$.

b) Pro náhodný jev $A \subseteq \bigcup_{i=1}^{n} B_i$, kde B_i jsou disjunktní náhodné jevy, $i = 1, \ldots, n$, je tzv. ***úplná pravděpodobnost***

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A \mid B_i)$$

a pro $P(A) \neq 0$ platí ***Bayesův vzorec***

$$P(B_j \mid A) = \frac{P(B_j)P(A \mid B_j)}{\sum_{i=1}^{n} P(B_i)P(A \mid B_i)}, \quad j = 1, \ldots, n\,.$$

Předpoklad $A \subseteq \bigcup_{i=1}^{n} B_i$ ve vlastnosti b) se často nahrazuje předpokladem $\bigcup_{i=1}^{n} B_i = \Omega$ a hovoří se o ***úplné skupině disjunktních jevů*** B_i.

Příklad 4

Ze skupiny 100 výrobků, která obsahuje 10 zmetků, vybereme náhodně bez vracení 3 výrobky. Pravděpodobnost toho, že první výrobek není zmetek – náhodný jev A_1, druhý výrobek není zmetek – náhodný jev A_2 a třetí výrobek je zmetek – náhodný jev $\bar{A}_3$, je

$$P(A_1 \cap A_2 \cap \bar{A}_3) = P(A_1)P(A_2 \mid A_1)P(\bar{A}_3 \mid A_1 \cap A_2) = \frac{90}{100}\frac{89}{99}\frac{10}{98} \doteq 0{,}08256\,.$$

Příklad 5

Do obchodu s potravinami dodávají rohlíky stejného druhu 3 pekárny v počtech 500, 1000 a 1500 kusů denně. Zmetkovitost jejich dodávek je 5 %, 4 % a 3 %. Tyto dodávky jsou v obchodě smíchány do celkové zásoby. Určete pravděpodobnost, že

a) náhodně vybraný rohlík z celkové zásoby je zmetek,
b) tento rohlík byl dodán druhou pekárnou.

Ř e š e n í:

Označme náhodné jevy

A ... vybraný rohlík je zmetek,
B_i ... vybraný rohlík byl dodán i-tou pekárnou, $i = 1, 2, 3$.

Pravděpodobnosti jsou

$$P(B_1) = \frac{500}{500+1000+1500} = \frac{1}{6}, \quad P(A \mid B_1) = 0{,}05,$$
$$P(B_2) = \frac{1000}{500+1000+1500} = \frac{2}{6}, \quad P(A \mid B_2) = 0{,}04,$$
$$P(B_3) = \frac{1500}{500+1000+1500} = \frac{3}{6}, \quad P(A \mid B_3) = 0{,}03\,.$$

a) Podle vzorce pro úplnou pravděpodobnost je

$$P(A) = 0{,}05\frac{1}{6} + 0{,}04\frac{2}{6} + 0{,}03\frac{3}{6} = \frac{0{,}22}{6} = 0{,}03\overline{6} \doteq 0{,}03667\,,$$

takže zmetkovitost z hlediska zákazníka je přibližně 3,667%.

b) Z Bayesova vzorce pro $j = 2$ je

$$P(B_2 \mid A) = \frac{0{,}04\frac{2}{6}}{\frac{0{,}22}{6}} = \frac{0{,}08}{0{,}22} = 0{,}\overline{36} \doteq 0{,}36364\,.$$

Analogicky lze získat $P(B_1/A) \doteq 0{,}22727$ a $P(B_3/A) \doteq 0{,}40909$, takže největší podíl na zmetkovitosti celkové zásoby má 3. pekárna. Má sice absolutně nejmenší zmetkovitost ze všech tří dodavatelů, avšak dodává největší počet rohlíků.

jevy

Náhodné jevy A, $B \in \Sigma$ jsou ***nezávislé***, jestliže $P(A \mid B) = P(A)$ anebo $P(B) = 0$. Náhodné jevy $A_1, \ldots, A_n \in \Sigma$ jsou ***vzájemně nezávislé***, jestliže jsou nezávislé všechny dvojice náhodných jevů

A_i, A_j pro všechny indexy $i \neq j$,
$A_i, A_j \cap A_k$ pro všechny indexy $i \neq j,\ i \neq k$,
$A_i, A_j \cap A_k \cap A_m$ pro všechny indexy $i \neq j,\ i \neq k$ a $i \neq m$,
atd.

Pro nezávislé náhodné jevy platí:

a) A, B jsou nezávislé, právě když $P(A \cap B) = P(A)P(B)$.

b) Jestliže $A_1, \ldots, A_n$ jsou vzájemně nezávislé, pak:
- $P\left(A_1 \cap \ldots \cap A_n\right) = P\left(A_1\right) \cdots P\left(A_n\right)$,
- $P(A_1 \cup \ldots \cup A_n) = 1 - \left[1 - P(A_1)\right] \cdots \left[1 - P(A_n)\right]$,
- náhodné jevy $B_1, \ldots, B_n$ jsou vzájemně nezávislé pro libovolné varianty $B_i = A_i,\ \overline{A}_i,\ \Omega$.

Příklad 6
Jaká je pravděpodobnost, že v prvním hodu pravidelnou homogenní šestistěnnou kostkou padne sudé číslo (náhodný jev A) a ve druhém hodu touto kostkou padne liché číslo (náhodný jev B), jestliže se kostka při hodech nedeformuje?
Ř e š e n í:
Náhodné jevy A a B jsou nezávislé a jejich pravděpodobnosti jsou
$P(A) = P(B) = 1/2$, takže $P(A \cap B) = (1/2)\cdot(1/2) = 1/4$.

Příklad 7
Vypočtěte řešený příklad 2.4 pro případ, že po každém výběru vracíme vybraný výrobek zpět.
Ř e š e n í:
Jevy A_1, A_2 a $\overline{A}_3$ jsou vzájemně nezávislé, takže

$$P(A_1 \cap A_2 \cap \overline{A}_3) = P(A_1)P(A_2)P(\overline{A}_3) = \frac{90}{100}\frac{90}{100}\frac{10}{100} = 0,081.$$

Příklad 8
Výrobek prochází třemi nezávislými operacemi, při kterých jsou pravděpodobnosti výroby zmetku $P(A_1) = 0,05$, $P(A_2) = 0,08$ a $P(A_3) = 0,03$. Určete pravděpodobnost výroby zmetku po všech třech operacích.
Ř e š e n í:
Vzhledem k nezávislosti operací jsou vzájemně nezávislé i náhodné jevy A_1, A_2, A_3 a výrobek je zmetek, jestliže nastane aspoň jeden z těchto jevů, takže

$$P(A_1 \cup A_2 \cup A_3) = 1 - \left[1 - P(A_1)\right]\left[1 - P(A_2)\right]\left[1 - P(A_3)\right] =$$
$$= 1 - 0,95 \cdot 0,92 \cdot 0,97 = 0,15222$$

Náhodná veličina a její charakteristiky

Náhodná veličina a distribuční funkce

Náhodná veličina (***náhodná proměnná***) X je reálná proměnná, která nabývá náhodně reálných číselných hodnot x. Uvedené vymezení pojmu náhodné veličiny má pouze intuitivní charakter. Matematickou definici lze najít např. v [2], [3], [5], [8, [9], [13], [14]. Množina všech hodnot náhodné veličiny X se nazývá ***základní soubor*** nebo také ***populace***. ***Distribuční funkce*** náhodné veličiny X je reálná funkce

$$F(x) = P(X < x) = P(X \in (-\infty; x)),$$

definovaná pro všechna $x \in (-\infty; +\infty)$. Distribuční funkce má vlastnosti:

a) $0 \le F(x) \le 1$ pro všechna $x \in (-\infty;+\infty)$,

b) $F(x)$ je neklesající, zleva spojitá a má nejvýše spočetně mnoho bodů nespojitosti na $(-\infty;+\infty)$,

c) $\lim\limits_{x\to-\infty} F(x) = F(-\infty) = 0$, $\lim\limits_{x\to+\infty} F(x) = F(+\infty) = 1$,

d) $P(a \le X < b) = F(b) - F(a)$ pro libovolná reálná čísla $a < b$,
speciálně $P(a \le X) = 1 - F(a)$, $P(X < b) = F(b)$,
$P(-\infty < X) = P(X < \infty) = 1$,

e) $P(X = c) = \lim\limits_{x\to c+} F(x) - F(c)$ pro libovolné reálné číslo c.

Distribuční funkcí je náhodná veličina X plně popsána a říkáme, že je dáno její ***rozdělení pravděpodobnosti***.

Někdy se distribuční funkce definuje vztahem $F(x) = P(X \le x)$. Tato distribuční funkce je však zprava spojitá, $P(a < X \le b) = F(b) - F(a)$ a $P(X = c) = F(c) - \lim\limits_{x\to c-} F(x)$. Setkáváme se s ní obvykle ve statistických softwarových produktech.

Diskrétní náhodná veličina

Náhodná veličina X je ***diskrétní*** a říkáme, že má ***diskrétní rozdělení pravděpodobnosti***, jestliže nabývá nejvýše spočetně mnoha hodnot

$$x = x_1, x_2, \dots .$$

Její ***pravděpodobnostní funkce*** je posloupnost

$$p(x) = P(X = x) > 0 \quad \text{pro } x = x_1, x_2, \dots .$$

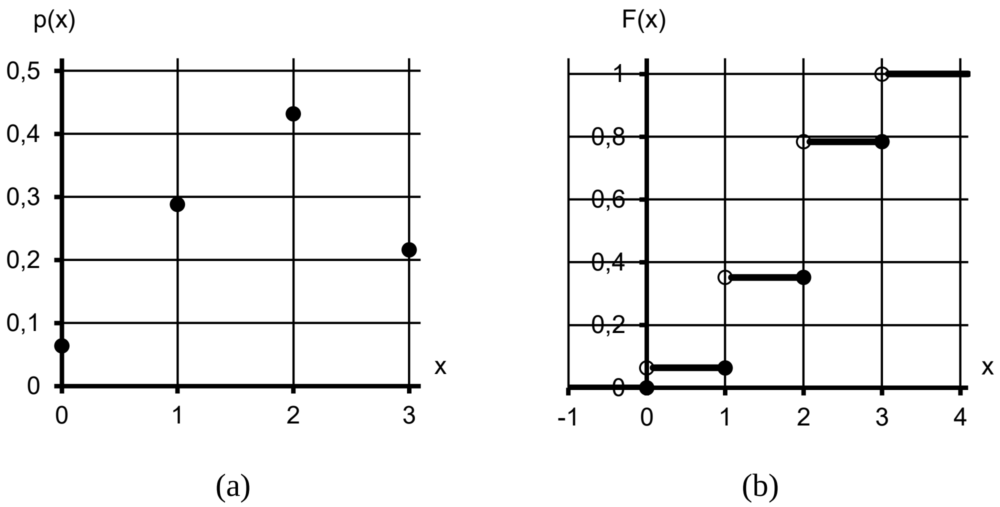
Pravděpodobnostní funkce má vlastnosti:

a) $\sum\limits_{x} p(x) = 1$,

b) $F(x) = \sum\limits_{t<x} p(t)$ pro všechna $x \in (-\infty;+\infty)$,

c) $P(X \in M) = \sum\limits_{x\in M} p(x)$ pro libovolnou množinu reálných čísel M.

Distribuční funkce diskrétní náhodné veličiny má "schodovitý tvar" – viz obr. 2. Z vlastnosti b) je zřejmé, že pravděpodobnostní funkcí je náhodná veličina jednoznačně určena.



Obr. 2

Příklad 9

Pravděpodobnost poruchy každé ze tří nezávisle pracujících výrobních linek je p, $0 < p < 1$. Diskrétní náhodná veličina X, která vyjadřuje počet výrobních linek v poruše, nabývá hodnot $x = 0, 1, 2, 3$ a hodnoty její pravděpodobnostní funkce jsou

$$p(0) = (1-p)^3,$$
$$p(1) = 3p(1-p)^2,$$
$$p(2) = 3p^2(1-p),$$
$$p(3) = p^3.$$

Dle potřeby zapisujeme pravděpodobnostní funkci do tabulky:

x	0	1	2	3
$p(x)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

Její distribuční funkce je

$F(x) = 0$ pro $x \in (-\infty, 0\rangle$,
$F(x) = p(0) = (1-p)^3$ pro $x \in (0, 1\rangle$,
$F(x) = p(0) + p(1) = (1 + 2p)(1-p)^2$ pro $x \in (1, 2\rangle$,
$F(x) = p(0) + p(1) + p(2) = (1 + p + p^2)(1-p) = 1 - p^3$ pro $x \in (2, 3\rangle$,
$F(x) = p(0) + p(1) + p(2) + p(3) = 1$ pro $x \in (3; +\infty)$.

Na obr. 2 jsou grafy p(x) a F(x) pro $p = 0{,}6$. Pravděpodobnost toho, že alespoň jedna linka má poruchu je

$$P(X \geq 1) = P(1 \leq X < +\infty) = F(+\infty) - F(1) = 1 - (1-p)^3.$$

Spojitá náhodná veličina

Náhodná veličina X je ***spojitá*** a říkáme, že má ***spojité rozdělení pravděpodobnosti***, jestliže má spojitou distribuční funkci $F(x)$ pro všechna $x \in (-\infty;+\infty)$. Její ***hustota pravděpodobnosti*** je taková nezáporná funkce $f(x)$, že pro všechna $x \in (-\infty;+\infty)$ je

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t\,.$$

Hustota pravděpodobnosti má vlastnosti:

a) $\displaystyle\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$,

b) $f(x) = F'(x)$, pokud derivace existuje,

c) $P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) =$
$$= \int_a^b f(x)\mathrm{d}x = F(b) - F(a) \quad \text{pro libovolná reálná čísla } a \leq b,$$

d) $P(X = c) = 0$ pro libovolné reálné číslo c.

Spojitá náhodná veličina obvykle nabývá všech hodnot z nějakého intervalu a je až na zanedbatelnou množinu hodnot (přesněji množinu nulové míry) jednoznačně určena hustotou pravděpodobnosti.

Příklad 10

Náhodná veličina X má hustotu pravděpodobnosti $f(x) = cx$ pro $x \in \langle 0; 2\rangle$ a 0 pro $x \notin \langle 0; 2\rangle$. Z vlastností spojité náhodné veličiny získáme následující výsledky. Je

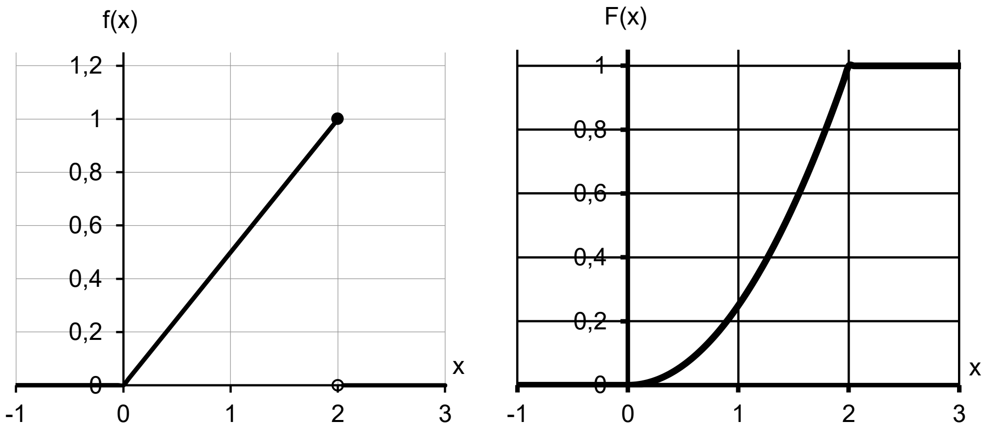
$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_{-\infty}^{0} 0\mathrm{d}x + \int_{0}^{2} cx\mathrm{d}x + \int_{2}^{+\infty} 0\mathrm{d}x = \ldots = 2c = 1,$$

takže $c = 1/2$ a $f(x) \geq 0$ pro všechna $x \in (-\infty;+\infty)$. Distribuční funkce náhodné veličiny X je

$$F(x) = \int_{-\infty}^{x} 0\mathrm{d}t = 0 \quad \text{pro } x \in (-\infty;\, 0\rangle,$$

$$F(x) = \int_{-\infty}^{0} 0\mathrm{d}t + \int_{0}^{x} \frac{t}{2}\mathrm{d}t = \ldots = \frac{x^2}{4} \quad \text{pro } x \in \langle 0;\, 2\rangle,$$

$$F(x) = \int_{-\infty}^{0} 0\mathrm{d}t + \int_{0}^{2} \frac{t}{2}\mathrm{d}t + \int_{2}^{x} 0\mathrm{d}t = \ldots = 1 \quad \text{pro } x \in \langle 2;+\infty).$$



(a) (b)

Obr. 3

Na obr. 3 jsou grafy $f(x)$ a $F(x)$. Pravděpodobnost toho, že náhodná veličina nabude hodnotu $x \in \langle 1;\, 3\rangle$, je $P(1 \leq X \leq 3) = F(3) - F(1) = 1 - (1^2/4) = 0{,}75$.

Číselné charakteristiky náhodné veličiny

Číselné charakteristiky náhodné veličiny X jsou reálná čísla, která v koncentrované formě vyjadřují její důležité vlastnosti.

Polohu rozdělení pravděpodobnosti náhodné veličiny X charakterizuje její ***střední hodnota***

$E(X) = \sum_{x} xp(x)$ pro diskrétní náhodnou veličinu X (pokud řada konverguje absolutně),

$E(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$ pro spojitou náhodnou veličinu X (pokud integrál konverguje absolutně).

Střední hodnota má vlastnosti:

a) $E(aX + b) = aE(X) + b$ pro libovolná reálná čísla a, b,

b) $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$ pro náhodné veličiny $X_1, \ldots, X_n$,

c) $E(X)$ má tentýž rozměr jako náhodná veličina X.

Míru kolísání hodnot náhodné veličiny X kolem její střední hodnoty $E(X)$ vyjadřuje její ***rozptyl*** (***disperze***, ***variance***) $D(X) = E([X - E(X)]^2)$. Rozptyl má vlastnosti:

a) $D(X) = \sum_x (x - E(X))^2 p(x) = \sum_x x^2 p(x) - (E(X))^2$ pro diskrétní náhodnou veličinu X (pokud řada konverguje),

b) $D(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) \mathrm{d}x = \int_{-\infty}^{+\infty} x^2 f(x) \mathrm{d}x - (E(X))^2$ pro spojitou náhodnou veličinu X (pokud integrál konverguje),

c) $D(X) \geq 0$,

d) $D(aX + b) = a^2 D(X)$ pro libovolná reálná čísla a, b,

e) $D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i)$ pro nezávislé náhodné veličiny $X_1, \ldots, X_n$,

f) $D(X)$ má rozměr rovný kvadrátu rozměru náhodné veličiny X.

Většímu rozptylu $D(X)$ odpovídá větší rozptyl pozorovaných hodnot náhodné veličiny X a naopak.

Směrodatná odchylka náhodné veličiny X je $\sigma(X) = \sqrt{D(X)}$.

Směrodatná odchylka má vlastnosti:

a) $\sigma(X) \geq 0$,

b) $\sigma(aX + b) = |a|\, \sigma(X)$ pro libovolná reálná čísla a, b,

c) $\sigma(X)$ má tentýž rozměr jako náhodná veličina X.

Větší směrodatné odchylce $\sigma(X)$ odpovídá větší rozptyl pozorovaných hodnot náhodné veličiny X a naopak.

Střední hodnota, popř. rozptyl, náhodné veličiny X je speciální případ tzv. ***obecného***, popř. ***centrálního momentu***. Blíže o ***momentových charakteristikách*** (***variačním koeficientu***, ***koeficientech šikmosti*** a ***špičatosti***) najdete v [1], [2], [3].

Momentové charakteristiky náhodné veličiny X nemusí vždy z důvodu požadavku konvergence existovat, avšak některé jiné číselné charakteristiky lze vždy určit. Jde především o tzv. ***P-kvantil*** nebo také **100*P* %-*kvantil*** náhodné veličiny X. Je jím pro $0 < P < 1$ její hodnota $x_P = \inf \{x; F(x) \geq P\}$. Pro spojitou náhodnou veličinu X s rostoucí distribuční funkcí je $F(x_P) = P$. Kvantil $x_{0,5}$ je ***medián*** náhodné veličiny X a charakterizuje polohu jejího rozdělení pravděpodobnosti. Další ***kvantilové charakteristiky*** jsou uvedeny např. v [8], [14], [17], [20], [30].

Modus $\hat{x}$ náhodné veličiny X je její hodnota, v níž nabývá pravděpodobnostní funkce nebo hustota pravděpodobnosti maximum, příp. suprémum.

Příklad 11

Náhodná veličina X z příkladu 9 má střední hodnotu

$$E(X) = \sum_{x=0}^{3} x p(x) = \cdots = 3p,$$

rozptyl

$$D(X) = \sum_{x=0}^{3} x^2 p(x) - (3p)^2 = \cdots = 3p(1-p)$$

a směrodatnou odchylku

$$\sigma(X) = \sqrt{3p(1-p)}.$$

Tyto charakteristiky (a také další) dané náhodné veličiny X je možno bez výpočtu určit přímo ze vztahů z kapitoly 4, neboť X má tzv. ***binomické rozdělení pravděpodobnosti***.

Příklad 12

Náhodná veličina X z příkladu 10 má střední hodnotu

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \int_{-\infty}^{0} x0\mathrm{d}x + \int_{0}^{2} x\frac{x}{2}\mathrm{d}x + \int_{2}^{+\infty} x0\mathrm{d}x = \cdots = \frac{4}{3} \doteq 1{,}33333,$$

rozptyl

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x - (E(X))^2 = \int_{-\infty}^{0} x^2 0\mathrm{d}x + \int_{0}^{2} x^2\frac{x}{2}\mathrm{d}x + \int_{2}^{+\infty} x^2 0\mathrm{d}x - \left(\frac{4}{3}\right)^2 =$$

$$= 2 - \frac{16}{9} = \frac{2}{9} \doteq 0{,}22222$$

a směrodatnou odchylku

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{2}{9}} \doteq 0{,}47140.$$

P-kvantil x_P je kořen rovnice $F(x_P) = P$, takže hledáme řešení rovnice $\frac{x^2}{4} = P$ z intervalu $\langle 0;$ $2\rangle$, jímž je $x_P = 2\sqrt{P}$. Odtud medián náhodné veličiny X je $x_{0,5} = 2\sqrt{0{,}5} \approx 1{,}41421$. Z grafu $f(x)$ na obr. 3 vidíme, že modus náhodné veličiny X je $\hat{x} = 2$.

Diskrétní rozdělení pravděpodobnosti

a) ***Binomické rozdělení*** Bi(n, p), kde n je přirozené číslo, p je reálné číslo, $0 < p < 1$:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n;$$

$$E(X) = np; \quad D(X) = np(1-p); \quad (n+1)p - 1 \le \hat{x} \le (n+1)p.$$

Toto rozdělení má počet x nastoupení sledovaného náhodného jevu v posloupnosti n vzájemně nezávislých pokusů. Jedná se také o popis tzv. ***náhodného výběru s vracením***, kdy např. postupně vybíráme z dodávky n výrobků ke kontrole, x je počet zmetků mezi nimi, p je pravděpodobnost výroby zmetku, a každý vybraný výrobek vracíme zpět do dodávky (to znamená, že může být znovu kontrolován).

Příklad 13

V dodávce 50 výrobků je 5 zmetků. Z dodávky jsou náhodně vybrány 3 výrobky. Počet zmetků mezi vybranými výrobky je náhodná veličina X. Určete typ jejího rozdělení pravděpodobnosti, její pravděpodobnostní funkci $p(x)$, střední hodnotu $E(X)$, rozptyl $D(X)$, směrodatnou odchylku $\sigma(X)$, medián $x_{0,5}$, modus $\hat{x}$ a $P(1 < X \le 3)$. Předpokládejte, že každý vybraný výrobek se vrátí nazpět do dodávky, takže jde o náhodný výběr s vracením.

Ř e š e n í:

Náhodná veličina X má rozdělení Bi(n,p), kde $n = 3$ a $p = 5/50 = 0{,}1$. Náhodná veličina X nabývá hodnot $x = 0, 1, 2, 3$ a její pravděpodobnostní funkce je

$$p(x) = \binom{3}{x} 0{,}1^x \cdot 0{,}9^{3-x} \text{ pro } x = 0, 1, 2, 3.$$

Střední hodnota je $E(X) = np = 3 \cdot 0{,}1 = 0{,}3$,

rozptyl je $D(X) = np(1-p) = 3 \cdot 0{,}1 \cdot 0{,}9 = 0{,}27$, směrodatná odchylka je $\sigma(X) = \sqrt{D(X)} = \sqrt{0,27} \doteq 0,51962$,

medián $x_{0,5} = 0$, neboť $F(x) = p(0) = \binom{3}{0} 0,1^0 \cdot 0,9^{3-0} = 0,729 > 0,5$, $x \in (0;1\rangle$,

modus $\hat{x} = 0$, neboť $(n + 1)p - 1 = -0,6$ a $(n + 1)p = 0,4$,
$P(1 < X \leq 3) = p(2) + p(3) = 0,027 + 0,001 = 0,028$.

b) ***Hypergeometrické rozdělení*** H(N, M, n), kde N, M a n jsou přirozená čísla, $1 \leq n \leq N$, $1 \leq M \leq N$:

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad x = \max\{0, M - N + n\}, \ldots, \min\{M, N\};$$

$$E(X) = n\frac{M}{N}; \quad D(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}; \quad a - 1 \leq \hat{x} \leq a, \text{ kde } a = \frac{(M+1)(n+1)}{N+2}.$$

Toto rozdělení popisuje tzv. ***náhodný výběr bez vracení***, kdy např. N je celkový počet výrobků, M je počet zmetků mezi těmito výrobky a vybereme náhodně (bez vracení jednotlivých výrobků nebo jejich skupin) celkem n výrobků, mezi nimiž je x zmetků.

Příklad 14

V dodávce 50 výrobků je 5 zmetků. Z dodávky jsou náhodně vybrány 3 výrobky. Počet zmetků mezi vybranými výrobky je náhodná veličina X. Určete typ jejího rozdělení pravděpodobnosti, její pravděpodobnostní funkci $p(x)$, střední hodnotu $E(X)$, rozptyl $D(X)$, směrodatnou odchylku $\sigma(X)$, medián $x_{0,5}$, modus $\hat{x}$ a $P(1 < X \leq 3)$. Předpokládejte (na rozdíl od řešeného příkladu 13), že se vybraný výrobek nevrací nazpět do dodávky, takže jde o náhodný výběr bez vracení.
Ř e š e n í:
Náhodná veličina X má rozdělení H(N, M, n), kde $N = 50$, $M = 5$ a $n = 3$. Náhodná veličina X nabývá hodnot $x = 0, 1, 2, 3$ a její pravděpodobnostní funkce je

$$p(x) = \frac{\binom{5}{x}\binom{45}{3-x}}{\binom{50}{3}} \quad \text{pro } x = 0, 1, 2, 3.$$

Střední hodnota je $E(X) = n\frac{M}{N} = 3\cdot 0,1 = 0,3$,

rozptyl je $D(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1} = 3\cdot 0,1\cdot 0,9\cdot (47/49) \doteq 0,25898$,

směrodatná odchylka je $\sigma(X) = \sqrt{D(X)} \doteq \sqrt{0,25898} \doteq 0,50890$,

medián $x_{0,5} = 0$, neboť $F(x) = p(0) = \frac{\binom{5}{0}\binom{45}{3-0}}{\binom{50}{3}} \doteq 0,72398 > 0,5$ pro $x(0;1\rangle$,

modus $\hat{x} = 0$, neboť $a = \frac{(M+1)(n+1)}{N+2} \doteq 0,46154$, takže $a - 1 \doteq -0,53846$,
$P(1 < X \leq 3) = p(2) + p(3) \doteq 0,02296 + 0,00051 = 0,02347$.

c) ***Poissonovo rozdělení*** Po(λ), kde λ je reálné číslo, $\lambda > 0$:

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, \dots ;$$

$$E(X) = \lambda; \;\; D(X) = \lambda; \;\; \lambda - 1 \le \hat{x} \le \lambda.$$

Toto rozdělení se obvykle užívá pro vyjádření pravděpodobnosti počtu nastoupení sledovaného jevu v určitém časovém intervalu (počet poruch, nehod, katastrof, zmetků apod.) s malou pravděpodobností výskytu.

Příklad 15

Statistickým průzkumem bylo zjištěno, že během jedné minuty navštíví prodejnu průměrně 3 zákazníci. Najděte vhodný typ rozdělení pravděpodobnosti náhodné veličiny X vyjadřující počet zákazníků, kteří navštíví prodejnu během jedné minuty. Určete její pravděpodobnostní funkci $p(x)$, střední počet zákazníků $E(X)$, rozptyl $D(X)$ a směrodatnou odchylku $\sigma(X)$ počtu zákazníkům a nejpravděpodobnější počet zákazníků za jednu minutu. Určete dále pravděpodobnost, že během jedné minuty přijde (a) právě 1 zákazník, (b) aspoň 1 zákazník, (c) medián $x_{0,5}$ počtu zákazníků.

Ř e š e n í:

Nahradíme střední počet zákazníků, kteří navštíví prodejnu během jedné minuty, jejich průměrným počtem, tj. položíme $E(X) = \bar{x}$. Vzhledem k tomu, že nemáme další informace o rozdělení pravděpodobnosti náhodné veličiny X (např. o rozptylu $D(X)$), použijeme Poissonovo rozdělení pravděpodobnosti Po(λ) s pravděpodobnostní funkcí

$$p(x) = \frac{3^x}{x!} e^{-3}, \quad x = 0, 1, \dots$$

Střední hodnota je $E(X) = \lambda = 3$,

rozptyl je $D(X) = \lambda = 3$,

směrodatná odchylka je $\sigma(X) = \sqrt{D(X)} = \sqrt{3} \doteq 1,73205$,

pro modus (nejpravděpodobnější počet zákazníků) je $\lambda - 1 \le \hat{x} \le \lambda$, takže $\hat{x} = 2$ a 3.

(a) $P(X = 1) = p(1) = \dfrac{3^1}{1!} e^{-3} \doteq 0,14936$,

(b) $P(X \ge 1) = p(1) + p(2) + \dots = 1 - p(0) = 1 - \dfrac{3^0}{0!} e^{-3} \doteq 1 - 0,04979 =$
$= 0,95021$,

(c) medián je $x_{0,5} = 3$, neboť $p(0) + p(1) + p(2) \doteq 0,42319 < 0,5$
a $p(0) + p(1) + p(2) + p(3) \doteq 0,64723 > 0,5$.

Spojitá rozdělení pravděpodobnosti

a) ***Rovnoměrné rozdělení*** R(a, b), kde a, b jsou reálná čísla, $a < b$:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{pro } x \in \langle a; b \rangle, \\ 0 & \text{pro } x \notin \langle a; b \rangle; \end{cases}$$

$$F(x)=\begin{cases} 0 & \text{pro} \quad x\in(-\infty;a), \\ \dfrac{x-a}{b-a} & \text{pro} \quad x\in\langle a;b\rangle, \\ 1 & \text{pro} \quad x\in(b;+\infty); \end{cases}$$

$$E(X)=x_{0,5}=\frac{a+b}{2}; \quad D(X)=\frac{(b-a)^2}{12}; \quad A(X)=0.$$

Toto rozdělení slouží především k simulaci reálných procesů nebo numerickým výpočtům tzv. ***metodou Monte Carlo*** na počítači pomocí generátorů tzv. ***pseudonáhodných čísel***.

Příklad 16

K přerušení optického kabelu v délce 500 m může dojít v libovolné vzdálenosti od jeho počátku, přičemž pravděpodobnost náhodného jevu, že dojde k přerušení v nějakém úseku je přímo úměrná délce úseku a nezávisí na jeho poloze. Určete rozdělení pravděpodobnosti náhodné veličiny X vyjadřující vzdálenost místa přerušení od počátku, její hustotu pravděpodobnosti a základní číselné charakteristiky a pravděpodobnost, že k přerušení kabelu dojde v úseku od 300 m do 400 m.

Ř e š e n í:

Náhodná veličina X má rozdělení R(a, b), kde $a = 0$ a $b = 500$ s hustotou pravděpodobnosti

$$f(x)=\begin{cases} \dfrac{1}{500} & \text{pro} \quad x\in\langle 0;500\rangle, \\ 0 & \text{pro} \quad x\notin\langle 0;500\rangle. \end{cases}$$

Střední hodnota a medián vzdálenosti je $E(X)=x_{0,5}=\dfrac{0+500}{2}$ = 250 m,

rozptyl je $D(X)=\dfrac{(500-0)^2}{12} \doteq 20833{,}3\ \text{m}^2$,

směrodatná odchylka je $\sigma(X) = \sqrt{D(X)} \doteq \sqrt{20833{,}3} \doteq 144{,}34$ m,

pravděpodobnost $P(300 \le X \le 400) = F(400) - F(300) = \dfrac{400}{500}-\dfrac{300}{500} = 0{,}2$.

(b) ***Normální rozdělení*** N(μ, σ^2), kde μ, σ^2 jsou reálná čísla, $\sigma^2 > 0$:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in (-\infty, +\infty);$$

$$E(X) = x_{0,5} = \hat{x} = \mu; \quad D(X) = \sigma^2; \quad A(X) = 0.$$

Toto nejčastěji užívané rozdělení, nazývané také ***Gaussovo rozdělení***, má řadu významných teoretických vlastností a z hlediska aplikací bývá vhodné k vyjádření náhodných veličin, které lze interpretovat jako aditivní výsledek mnoha nezávislých vlivů (např. chyba měření, odchylka rozměru výrobku od požadované hodnoty apod.). Někdy se také hovoří o ***zákonu chyb***.

Transformací náhodné veličiny X s normálním rozdělením N(μ, σ^2) na náhodnou veličinu

$$U=\frac{X-\mu}{\sigma}$$

dostaneme ***normované*** (***základní***) ***normální rozdělení*** N(0;1) s distribuční funkcí $\Phi(u)$, jehož distribuční funkce $\Phi(x)$ je tabelována (viz tabulku **T1**) anebo její hodnoty určíme výpočtem na PC, např. pomocí softwaru Excel. Pro hodnoty $\Phi(u)$ platí

$$\Phi(-u) = 1 - \Phi(u).$$

Jestliže náhodná veličina X má normální rozdělení N(μ, σ^2), potom její distribuční funkce

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

V aplikacích (při řízení jakosti výroby apod.) se často užívá tzv. ***pravidlo tří sigma***, založené na tom, že

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = F(\mu + 3\sigma) - F(\mu - 3\sigma) = \Phi(3) - \Phi(-3) \doteq 0{,}9973.$$

Toto pravidlo znamená, že při velkém počtu pozorování náhodné veličiny X s normálním rozdělením N(μ, σ^2) můžeme očekávat, že všech cca 99,7 % pozorovaných hodnot x bude ležet v intervalu $<\mu - 3\sigma;\ \mu + 3\sigma>$. V tabulce T1 jsou hodnoty distribuční funkce normovaného (základního) normálního rozdělení pravděpodobnosti, nazývaného také Gaussovo rozdělení pravděpodobnosti. Pro výpočty je možno také použít statistické vzorce z Excelu.

Příklad 17

Jaká je pravděpodobnost, že náhodná veličina X, která má rozdělení N(20;16), nabude hodnotu
a) menší než 16, b) větší než 20, c) v mezích od 12 do 28,
d) menší než 12 nebo větší než 28 ?

Ř e š e n í:

Ze vztahu $F(x) = \Phi\left(\frac{x-20}{4}\right)$ a tabulky **T1** dostaneme:

a) $P(X < 16) = F(16) = \Phi((16 - 20) / 4) = \Phi(-1) = 1 - \Phi(1) = 1 - 0{,}84135 = = 0{,}15865$
b) $P(X > 20) = 1 - P(X \le 20) = 1 - F(20) = 1 - \Phi((20 - 20) / 4) = 1 - \Phi(0) =$
$= 1 - 0{,}5 = 0{,}5$
c) $P(12 \le X \le 28) = F(28) - F(12) = \Phi((28 - 20) / 4) - \Phi((12 - 20) / 4) = = \Phi(2) - \Phi(-2) =$
$\Phi(2) - (1 - \Phi(2)) = 2\Phi(2) - 1 = 2\cdot 0{,}97725 - 1 = 0{,}95450$
d) $P((X < 12) \vee (X > 28)) = 1 - P(12 \le X \le 28) = 1 - 0{,}9545 = 0{,}0455.$

T1 Hodnoty distribuční funkce $\Phi(u)$ normovaného normálního rozdělení N(0;1)

u	0	1	2	3	4	5	6	7	8	9
0,0	0,50000	50399	50798	51197	51596	51994	52392	52791	53188	53586
0,1	53983	54380	54776	55172	55567	55962	56356	56750	57143	57535
0,2	57926	58317	58707	59096	59484	59871	60257	60642	61026	61409
0,3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0,4	65542	65910	66276	66640	67003	67365	67724	68082	68439	68793
0,5	69146	69498	69847	70195	70540	70884	71226	71566	71904	72241
0,6	72575	72907	73237	73565	73892	74216	74537	74857	75175	75490
0,7	75804	76115	76424	76731	77035	77337	77637	77935	78231	78524
0,8	78815	79103	79389	79673	79955	80234	80511	80785	81057	81327
0,9	81594	81859	82121	82382	82639	82894	83147	83398	83646	83891
1,0	84135	84375	84614	84850	85083	85314	85543	85769	85993	86214
1,1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1,2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1,3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774
1,4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1,5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1,6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1,7	95543	95637	95728	95819	95907	95994	96080	96164	96246	96327
1,8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1,9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670
2,0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2,1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574
2,2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
2,3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
2,4	99180	99202	99224	99245	99266	99286	99305	99324	99343	99361
2,5	99379	99396	99413	99430	99446	99461	99477	99492	99506	99520
2,6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2,7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
2,8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2,9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
3,0	99865	99869	99874	99878	99882	99886	99889	99893	99896	99900
3,1	99903	99906	99910	99913	99916	99918	99921	99924	99926	99929
3,2	99931	99934	99936	99938	99940	99942	99944	99946	99948	99950
3,3	99952	99953	99955	99957	99958	99960	99961	99962	99964	99965
3,4	99966	99968	99969	99970	99971	99972	99973	99974	99975	99976
3,5	99977	99978	99978	99979	99980	99981	99981	99982	99983	99983
3,6	99984	99985	99985	99986	99986	99987	99987	99988	99988	99989
3,7	99989	99990	99990	99990	99991	99991	99992	99992	99992	99992
3,8	99993	99993	99993	99994	99994	99994	99994	99995	99995	99995
3,9	99995	99995	99996	99996	99996	99996	99996	99996	99997	99997

4,00	99997	4,10	99998	4,20	99999	4,30	99999	4,40	99999	4,50	99999

Poznámka: $\Phi(-u) = 1 - \Phi(u)$; $u_{0,95} \approx 1{,}645$; $u_{0,975} \approx 1{,}960$; $u_{0,99} \approx 2{,}326$; $u_{0,995} \approx 2{,}576$

Kontrolní otázky

1. V čem spočívá náhodnost náhodného jevu? Uveďte konkrétní příklad.
2. Uveďte příklad, kdy nelze použít tzv. klasickou pravděpodobnost.
3. Aplikujte úplnou pravděpodobnost a Bayesův vzorec na problém z Vaší firmy nebo Vašeho okolí.
4. Uveďte konkrétní příklad na nezávislé náhodné jevy.
5. Vyjádřete $P(A \cup B)$ pro (a) nezávislé (b) “závislé” náhodné jevy.
6. Uveďte konkrétní příklady na diskrétní a spojité náhodné veličiny.
7. Jakými funkčními charakteristikami se popisuje diskrétní náhodná veličina?
8. Jakými funkčními charakteristikami se popisuje spojitá náhodná veličina?
9. Které číselné charakteristiky vyjadřují polohu rozdělení pravděpodobnosti?
10. Které číselné charakteristiky vyjadřují variabilitu rozdělení pravděpodobnosti?
11. Jaké základní vlastnosti má střední hodnota náhodné veličiny? Interpretujte je!
12. Jaké základní vlastnosti má rozptyl náhodné veličiny? Interpretujte je!
13. Určete střední hodnotu a medián ceny výrobku v €, jestliže je známa jeho cena v Kč.
14. Určete rozptyl a směrodatnou odchylku ceny výrobku v US $, jestliže je známa jeho cena v Kč.
15. Co vyjadřuje medián náhodné veličiny?
16. Co vyjadřuje modus náhodné veličiny?
17. Uveďte konkrétní příklad na binomické rozdělení pravděpodobnosti a popište význam jeho parametrů i číselných charakteristik.
18. Uveďte příklad na hypergeometrické rozdělení pravděpodobnosti a popište význam jeho parametrů i číselných charakteristik.
19. Uveďte konkrétní příklad na Poissonovo rozdělení pravděpodobnosti a popište význam jeho parametrů i číselných charakteristik.
20. Uveďte konkrétní příklad na normální rozdělení pravděpodobnosti a popište význam jeho parametrů i číselných charakteristik.