VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA
OSTRAVA

PRAVDĚPODOBNOST
A
STATISTIKA

Petr Otipka
Vladislav Šmajstrla

Vytvořeno v rámci projektu Operačního programu Rozvoje lidských zdrojů
CZ.04.1.03/3.2.15.1/0016
Studijní opory s převažujícími distančními prvky pro předměty teoretického základu studia.

Tento projekt je spolufinancován Evropským sociálním fondem
a státním rozpočtem České republiky

ESF – ROVNÉ PŘÍLEŽITOSTI PRO VŠECHNY

VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA
OSTRAVA

PRAVDĚPODOBNOST
A
STATISTIKA

Petr Otipka
Vladislav Šmajstrla

Vytvořeno v rámci projektu Operačního programu Rozvoje lidských zdrojů
CZ.04.1.03/3.2.15.1/0016
Studijní opory s převažujícími distančními prvky pro předměty teoretického základu studia.

Tento projekt je spolufinancován Evropským sociálním fondem
a státním rozpočtem České republiky

ESF – ROVNÉ PŘÍLEŽITOSTI PRO VŠECHNY

ISBN 80-248-1194-4

OBSAH

STUDIJNÍ OPORY S PŘEVAŽUJÍCÍMI DISTANČNÍMI PRVKY PRO PŘEDMĚTY TEORETICKÉHO ZÁKLADU STUDIA

je název projektu, který uspěl v rámci první výzvy Operačního programu Rozvoj lidských zdrojů. Projekt je spolufinancován státním rozpočtem ČR a Evropským sociálním fondem. Partnery projektu jsou Regionální středisko výchovy a vzdělávání, s.r.o. v Mostě, Univerzita obrany v Brně a Technická univerzita v Liberci. Projekt byl zahájen 5.1.2006 a bude ukončen 4.1.2008.

Cílem projektu je zpracování studijních materiálů z matematiky, deskriptivní geometrie, fyziky a chemie tak, aby umožnily především samostatné studium a tím minimalizovaly počet kontaktních hodin s učitelem. Je zřejmé, že vytvořené texty jsou určeny studentům všech forem studia. Studenti kombinované a distanční formy studia je využijí k samostudiu, studenti v prezenční formě si mohou doplnit získané vědomosti. Všem studentům texty pomohou při procvičení a ověření získaných vědomostí. Nezanedbatelným cílem projektu je umožnit zvýšení kvalifikace širokému spektru osob, které nemohly ve studiu na vysoké škole z různých důvodů (sociálních, rodinných, politických) pokračovat bezprostředně po maturitě.

V rámci projektu jsou vytvořeny jednak standardní učební texty v tištěné podobě, koncipované pro samostatné studium, jednak e-learningové studijní materiály, přístupné prostřednictvím internetu. Součástí výstupů je rovněž banka testových úloh pro jednotlivé předměty, na níž si studenti ověří, do jaké míry zvládli prostudované učivo.

Bližší informace o projektu můžete najít na adrese http://www.studopory.vsb.cz/.

Přejeme vám mnoho úspěchů při studiu a budeme mít radost, pokud vám předložený text pomůže při studiu a bude se vám líbit. Protože nikdo není neomylný, mohou se i v tomto textu objevit nejasnosti a chyby. Předem se za ně omlouváme a budeme vám vděčni, pokud nás na ně upozorníte.

ESF – ROVNÉ PŘÍLEŽITOSTI PRO VŠECHNY

ÚVOD

Tento distanční text je určen studentům VŠB-TU Ostrava.

Je členěn na dvě základní části. První z nich je věnována základům počtu pravděpodobnosti, druhá úvodu do problematiky matematické statistiky.

Autoři se zaměřili na srozumitelný výklad základních pojmů a na objasnění souvislostí mezi těmito pojmy. Důkazy vět omezili na důkazy základních vět a na takové, které ilustrují úvahy, vedoucí k těmto větám. Každá kapitola obsahuje příklady s podrobným řešením a v závěru sadu neřešených úloh s výsledky.

Kapitoly věnované základům počtu pravděpodobnosti jsou zaměřeny na definování pravděpodobnosti různými způsoby , na popis náhodné veličiny a náhodného vektoru. Jsou uvedeny důležité typy rozdělení pravděpodobnosti diskrétní i spojité náhodné veličiny.

Část věnovaná matematické statistice seznamuje s popisem statistických souborů, momentovými a kvantilovými charakteristikami, objasňuje pojmy lineární a nelineární regrese. Závěrečné kapitoly jsou věnovány statistické indukci – získávání odhadů parametrů základního souboru a testování statistických hypotéz.

Za cenné rady a připomínky k práci děkujeme Ivanu Kolomazníkovi a také recenzentům Jiřímu Vrbickému a Michalu Vavrošovi.

POKYNY KE STUDIU

V úvodu si vysvětlíme jednotnou pevnou strukturu každé kapitoly textu, která by vám měla pomoci k rychlejší orientaci při studiu. Pro zvýraznění jednotlivých částí textu jsou používány ikony a barevné odlišení, jejichž význam nyní objasníme.

Průvodce studiem

vás stručně seznámí s obsahem dané kapitoly a s její motivací. Slouží také k instrukci, jak pokračovat dál po vyřešení kontrolních otázek nebo kontrolních textů.

Cíle

vás seznámí s učivem, které v dané kapitole poznáte a které byste po jejím prostudování měli umět.

Předpokládané znalosti

shrnují stručně učivo, které byste měli znát ještě dříve než kapitolu začnete studovat. Jsou nezbytným předpokladem pro úspěšné zvládnutí následující kapitoly.

Výklad

označuje samotný výklad učiva dané kapitoly, který je členěn způsobem obvyklým v matematice na definice, věty, případně důkazy.

Definice 1.1.1.
Zavádí základní pojmy v dané kapitole.

Věta 1.1.1.
Uvádí základní vlastnosti pojmů zavedených v dané kapitole.

Důkaz: Vychází z předpokladů věty a dokazuje tvrzení uvedené ve větě.

Poznámka

neformálně komentuje vykládanou látku..

Řešené úlohy

označují vzorové příklady, které ilustrují probrané učivo.

Příklad Uvádí zadání příkladu.

Řešení: Uvádí podrobné řešení zadaného příkladu.

Úlohy k samostatnému řešení

obsahují zadání příkladů k procvičení probraného učiva. Úlohy označené ✖ patří k obtížnějším a jsou určeny zájemcům o hlubší pochopení tématu.

Výsledky úloh k samostatnému řešení

obsahují správné výsledky předchozích příkladů, slouží ke kontrole správnosti řešení.

Kontrolní otázky

obsahují soubor otázek k probranému učivu včetně několika odpovědí, z nichž je vždy alespoň jedna správná.

Odpovědi na kontrolní otázky

uvádějí správné odpovědi na kontrolní otázky.

Kontrolní test

obsahuje soubor příkladů k probranému učivu.

Výsledky testu

uvádějí správné odpovědi na příklady kontrolního testu.

Shrnutí lekce

obsahuje stručný přehled učiva, které by měl student po prostudování příslušné kapitoly zvládnout.

Literatura

obsahuje seznam knih, které byly použity při tvorbě příslušného textu a na které byly případně uvedeny odkazy k hlubšímu prostudování tématu.

Piktogram, který upozorňuje na důležité vztahy nebo vlastnosti, které je nezbytné si zapamatovat.

1. KOMBINATORIKA

Průvodce studiem

Na střední škole se někteří z vás seznámili se základními pojmy z kombinatoriky. V této kapitole tyto pojmy zopakujeme a prohloubíme vaše znalosti.

Předpokládané znalosti

Množiny. Faktoriál.

Cíle

Cílem této kapitoly je objasnit pojmy variace, permutace, kombinace.

Výklad

KOMBINATORIKA

Zkoumá skupiny (podmnožiny) prvků vybraných z jisté základní množiny. Podle toho, zda se prvky v jednotlivých skupinách **mohou či nemohou opakovat**, rozdělujeme skupiny prvků na **skupiny s opakováním a skupiny bez opakování.**

Poznámka

Skupiny, kde se prvky nemohou opakovat si lze tedy představit tak, že prvky, které vybíráme ze základní skupiny do ní nevracíme zpět a nemůžeme je tedy použít při dalším výběru. Naopak skupiny, kde se prvky mohou opakovat, vznikají tak, že vybrané prvky vracíme do základní skupiny a v dalším výběru je můžeme znovu použít.

Rozlišujeme tři základní způsoby výběru:

1.1. Variace *k*-té třídy z *n* prvků

- **uspořádané** skupiny po k prvcích z daných n prvků

Řešené úlohy

Příklad 1.1.1. Je dána množina M = {1,2,3,4,5}. Z prvků této množiny máme vytvářet dvojice, přičemž záleží na pořadí a prvky se nemohou opakovat.

Řešení: Vytváříme tedy variace druhé třídy z pěti prvků. Všechny možnosti:

$V_2(5)$: (1,2) (2,1) (1,3) (3,1) (1,4) (4,1) (1,5) (5,1)
(2,3) (3,2) (2,4) (4,2) (2,5) (5,2)
(3,4) (4,3) (3,5) (5,3)
(4,5) (5,4)

Takže počet všech možností je 20.

Příklad 1.1.2. Na startu běžeckého závodu je 8 atletů. Kolika způsoby mohou být obsazeny stupně vítězů?

Řešení: Jednoduchou úvahou dojdeme k tomu, že na prvním místě se může umístit kdokoliv z 8-mi startujících. Jestliže některý z atletů už doběhl první, druhé místo obsadí někdo ze zbývajících 7-mi závodníků. Jsou-li obsazena první dvě místa, je zřejmé, že pro třetí místo máme 6 možností.

Celkem tedy: $V_3(8) = 8.7.6 = 336$ možností

Obdobně můžeme postupovat při odvození obecného vzorce pro počet variací k-té třídy z n prvků bez opakování:

Ptáme se:

Z kolika prvků máme na výběr pro 1.člen k-tice?: n

Z kolika prvků máme na výběr pro 2.člen k-tice?: n - 1

. . .

Z kolika prvků máme na výběr pro k-tý člen k-tice?: n - k + 1

Proto:

$$V_k\;n\;= n.\;n-1\;\ldots\;n-k+1\;=$$

$$= n.\;n-1\;\ldots\;n-k+1\;.\frac{n-k\;.\;n-k-1\;\ldots 2.1}{n-k\;.\;n-k-1\;\ldots 2.1}=$$

$$=\frac{n!}{n-k\;!}$$

Takže:

1.1.1. Počet variací *k*-té třídy z *n* prvků bez opakování

$$V_k(n) = \frac{n!}{(n-k)!}$$

Řešené úlohy

Příklad 1.1.3. Kolik existuje trojciferných čísel, které lze zapsat užitím cifer 1, 2, 3, 4, 5.

Řešení: Jedná se o příklad na variace s opakováním - záleží na pořadí cifer a cifry se v čísle mohou opakovat:

Na první pozici v čísle se může vyskytovat libovolná cifra z daných pěti - tzn. 5 možností. Vzhledem k tomu, že cifry se v čísle mohou opakovat, dostáváme stejný počet možností i na druhé a třetí pozici. Počet všech možností:

$V_3^*(5) = 5.5.5 = 5^3 = 125$

Pokud tuto úvahu opět zobecníme dostaneme vzorec pro:

1.1.2. Počet variací *k*-té třídy z *n* prvků s opakováním

$$V_k^*(n) = n^k$$

Řešené úlohy

Příklad 1.1.4. Kolik různých značek teoreticky existuje v Morseově abecedě, sestavují-li se tečky a čárky do skupin po jedné až pěti?

Řešení: Máme k dispozici dva znaky: • –

Z těchto znaků vytváříme postupně jeden znak, dvojice, trojice, čtveřice a pětice. Záleží na pořadí, znaky se samozřejmě mohou opakovat, jedná se tedy o variace s opakováním, přičemž $n = 2$ a $k = 1, 2, 3, 4, 5$:

$z = V_1^*(2) + V_2^*(2) + V_3^*(2) + V_4^*(2) + V_5^*(2) = 2^1 + 2^2 + 2^3 + 2^4 + 2^5 =$

$= 2 + 4 + 8 + 16 + 32 = 62$

1.2. Permutace *n* prvků

- každá uspořádaná *n*-tice vybraná z *n* prvků

Řešené úlohy

Příklad 1.2.1. Najděte všechny permutace bez opakování z prvků množiny $M = \{1,7,9\}$

Řešení: Všechny permutace bez opakování z těchto tří prvků P(3):

(1,7,9), (1,9,7), (7,1,9), (7,9,1), (9,1,7), (9,7,1)

Příklad 1.2.2. Využijeme zadání příkladu 1.1.2., přičemž nás bude zajímat, kolika způsoby budou obsazena všechna místa.

Řešení: Vytváříme tedy osmice vybrané z osmi prvků, což přesně odpovídá pojmu permutace.

Úloha se dá vyřešit stejnou úvahou, jako příklad 1.1.2.. Na prvním místě máme 8 možností, na druhém 7 možností (první místo je již obsazeno), na třetím místě 6 možností, . . ., na osmém místě tedy zbývá pouze jediná možnost.

Výsledek je tedy $P(8) = 8.7.6.5.4.3.2.1 = 8! = 40320$ možností

Takže:

1.2.1. Počet permutací *n* prvků bez opakování

$$P(n) = n! = n.(n-1).(n-2)...3.2.1.$$

Řešené úlohy

Příklad 1.2.3. Mějme *n* různých korálků, které budeme navlékat na niť. Její konce pak svážeme, takže vytvoříme kruh (náhrdelník). Kolika způsoby lze korálky do kruhu uspořádat? Tzn. uspořádání, které se liší pouze otočením kruhu nepovažujeme za různé.

Řešení: Pokud bychom konce niti nesvázali, odpovídal by počet všech možností počtu permutací bez opakování z *n* prvků, těch je *n*! Ovšem v kruhu by některá z uspořádání byla shodná. Proveďme tedy následující úvahu. Uvažujme nějaké uspořádání v kruhu a zvolme si libovolný korálek, o kterém prohlásíme, že je první. Ostatní korálky očíslujeme např. ve směru hodinových ručiček. Celé uspořádání teď pootočíme ve směru hodinových ručiček o jeden korálek (první se dostane na místo druhého, druhý na místo třetího, ...), čímž v rámci kruhu dostaneme shodné uspořádání. Takto můžeme s korálky pootočit n krát a vždy dostaneme shodné uspořádání. Všechna tato shodná uspořádání jsou ale započítána do počtu $n!$ (počet uspořádání před svázáním konců niti). Výsledek je tedy:

$$x = \frac{n!}{n} = \frac{n.(n-1)!}{n} = (n-1)!$$

Příklad 1.2.4. Kolik různých šesticiferných čísel lze vytvořit z číslic 1, 2, 2, 3, 3, 3?

Řešení: Mezi danými šesti číslicemi se některé opakují. Pokud by se číslice neopakovaly, vytvořili bychom 6! čísel. V našem případě se počet čísel zmenší:

Z důvodu, že tam máme dvě dvojky se počet možností sníží dvakrát - jedna možnost 2 2 namísto dvou možností X 2, 2 X (permutace ze dvou prvků) v případě, že by číslice byly různé.

V důsledku tří trojek se počet čísel zmenší šestkrát - jedna možnost 3 3 3 namísto permutace ze tří různých číslic.

Počet všech možností je tedy:

$$P^*(6) = \frac{6!}{2!.3!}$$

Při zobecnění naší úvahy je:

1.2.2. Počet permutací *n* prvků s opakováním

$$P^*(n) = \frac{n!}{n_1!n_2!\ldots n_k!}$$

Jestliže se mezi n prvky vyskytuje: první prvek n_1 krát

druhý prvek n_2 krát $\Rightarrow n_1 + n_2 + \ldots + n_k = n$

…

k-tý prvek n_k krát

Řešené úlohy

Příklad 1.2.5. Zjistěte, kolik různých pěticiferných čísel lze vytvořit použitím cifer 1, 2, 3, 4, 5 (cifry se v čísle mohou opakovat).

Řešení: Při řešení této úlohy se často můžeme setkat s následující chybou: řešitel si všimne, že z pětiprvkové množiny máme vytvářet pětice a automaticky se úlohu snaží řešit pomocí permutací. Zde ale dochází ke kolizi, neboť o permutace bez opakování se jednat nemůže (cifry se v čísle mohou opakovat) a permutace s opakováním to být také nemohou (není určeno, kolikrát se který prvek má opakovat).

Zadání úlohy totiž přesně koresponduje s pojmem variace s opakováním, kde $k = n$, takže počet všech možností je:

$V_5^*(5) = 5^5 = 3125$

1.3. Kombinace *k*-té třídy z *n* prvků

- skupiny o k prvcích vybraných z n prvků

Poznámka

Vybíráme bez zřetele na uspořádání: tzn., že v daných n-ticích ***nezáleží*** *na pořadí prvků!*

Řešené úlohy

Příklad 1.3.1. Najděte všechny kombinace druhé třídy z množiny $M = \{1,2,3,4,5\}$

Řešení:

$C_2(5)$: (1,2) (1,3) (1,4) (1,5)
(2,3) (2,4) (2,5)
(3,4) (3,5)
(4,5)

Počet všech možností je tedy 10.

Příklad 1.3.2. Odvoďte počet kombinací *k*-té třídy z *n* prvků

Řešení: Umíme spočítat počet uspořádaných *k*-tic z *n* prvků - pomocí variací. Některé z těchto *k*-tic se však liší pouze pořadím prvků. Kolik jich je? Vezmeme libovolnou *k*-tici a vytvoříme všechny její obměny pouze s jejími prvky (tedy permutaci). Všechny *k*-tice, které jsme takto vytvořili, se budou lišit pouze pořadím prvků. Odtud je zřejmé, že počet kombinací *k*-té třídy z *n* prvků je:

$C_k(n) = V_k(n)/P(k)$:

1.3.1. Počet kombinací *k*-té třídy z *n* prvků bez opakování

$$C_k(n) = \frac{n!}{(n-k)!.k!} = \binom{n}{k}$$

Poznámka

$\binom{n}{k}$ ... kombinační číslo, čteme n nad k

Pro ruční výpočet kombinačních čísel je často vhodné použít následující odvození:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n.(n-1)\ldots(n-k+1).(n-k)!}{k!\,(n-k)!} = \frac{\overbrace{n.(n-1)\ldots(n-k+1)}^{k-\text{členů}}}{k!}$$

Takže například:

$$\binom{7}{3} = \frac{7.6.5.}{3.2.1} = 35$$

1.3.2. Počet kombinací *k*-té třídy z *n* prvků s opakováním

$$C_k^*(n) = \binom{n+k-1}{k}$$

Řešené úlohy

Příklad 1.3.3. Zjistěte, kolik existuje různých kvádrů, pro něž platí, že délka každé jejich hrany je přirozené číslo z intervalu $\langle 2,15 \rangle$

Řešení: Přirozených čísel v tomto intervalu je 14. Kvádr je jednoznačně určen třemi hodnotami (délka, šířka, výška) u nichž nezáleží na pořadí (je jedno, jak je kvádr "natočený"). Hodnoty v trojici se mohou opakovat (i krychle je speciální případ kvádru).

Takže se jedná o kombinace s opakováním, $n = 14$, $k = 3$:

$$C_3^*(14) = \binom{14+3-1}{3} = \binom{16}{3} = 560$$

1.3.3. Základní pravidla pro kombinační čísla

Symetrie

$$\binom{n}{k} = \binom{n}{n-k}$$

Okrajová vlastnost

$$\binom{n}{0} = \binom{n}{n} = 1$$

Sčítání

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

Řešené úlohy

Příklad 1.3.4. Řešte rovnici:

$$\binom{x+2}{x} + \binom{x+3}{x+1} = 64$$

Řešení:

$$\binom{x+2}{x} + \binom{x+3}{x+1} = 64$$

$$\binom{x+2}{2} + \binom{x+3}{2} = 64$$

$$\frac{x+2\ .\ x+1}{2.1} + \frac{x+3\ .\ x+2}{2.1} = 64$$

$$x^2 + 3x + 2 + x^2 + 5x + 6 = 128$$

$$2x^2 + 8x + 8 - 128 = 0$$

$$x^2 + 4x - 60 = 0$$

$$x+10\ .\ x-6\ = 0$$

$$x = 6$$

(kořen x = -10 nelze použít, x musí být přirozené číslo)

1.4. Řešené příklady, kombinatorika - souhrnně

Příklad 1.4.1. Jsou dány cifry 1, 2, 3, 4, 5. Cifry nelze opakovat. Kolik je možno vytvořit z těchto cifer čísel, která jsou:

a) pětimístná, sudá

b) pětimístná, končící dvojčíslím 21

c) pětimístná, menší než 30000

d) trojmístná lichá

e) čtyřmístná, větší než 2000

f) dvojmístná nebo trojmístná

Řešení:

ad *a*)

Sudá - to v tomto případě znamená, že končí ciframi 2 nebo 4 (XXXX2, XXXX4) - tzn. dvě možnosti. Na zbývajících čtyřech pozicích permutují zbývající čtyři cifry, takže výsledek:

$a = 2.P(4) = 48$

ad *b*)

Máme číslo XXX21. Tedy na třech pozicích permutují tři cifry:

$b = P(3) = 6$

ad *c*)

Menší než 30000, to jsou čísla začínající ciframi 1 nebo 2, tedy dvě možnosti. Na zbývajících čtyřech pozicích permutují zbývající čtyři cifry:

$c = 2.P(4) = 48$

ad *d*)

Lichá, tedy končí ciframi 1, 3, 5 - tři možnosti. Na zbývajících dvou pozicích se mohou vyskytovat některé ze zbývajících čtyř cifer, přičemž záleží na pořadí - jedná se o variace druhé třídy ze čtyř prvků.

$d = 3.V_2(4) = 36$

ad *e*)

obdobně jako u předchozích:

$e = 4.V_3(4) = 96$

ad *f*)

$f = V_2(5) + V_3(5) = 80$

Příklad 1.4.2. Kolik různých státních poznávacích značek OSB XX-XX existuje s aspoň dvěmi trojkami?

Řešení: Aspoň dvě trojky, to jsou 2, 3 nebo 4 trojky. Začneme nejjednodušší možností:

4 trojky:

Tzn. jediná možnost OSB 33-33, takže $x_4 = 1$

3 trojky:

Existují 4 možnosti, jak seskládat tři trojky na čtyřech pozicích (333X, 33X3, 3X33, X333). Obecně to lze vyjádřit jako počet permutací 4 prvků s opakováním, přičemž trojka se opakuje třikrát:

$$P^*(4) = \frac{4!}{3!} = 4$$

Dále existuje 9 možností (zbývajících devět cifer), které mohou být na čtvrté pozici. Obecně lze vyjádřit např. jako počet variací první třídy z devíti prvků:

$V_1(9) = 9$

Takže výsledný počet pro 3 trojky: $x_3 = P^*(4).V_1(9) = 4.9 = 36$

2 trojky:

Existuje opět $P^*(4)$ možností, jak seskládat dvě trojky na čtyři pozice, přičemž tentokrát se trojka opakuje dvakrát a zbývající dvě pozice nerozlišujeme mezi sebou, takže se také dvakrát opakují (33XX, 3X3X, 3XX3, X33X, X3X3, XX33):

$$P^*(4) = \frac{4!}{2!.2!} = 6$$

Na zbývajících dvou pozicích se může střídat zbývajících devět cifer, přičemž v dané dvojici záleží na pořadí cifer a cifry se mohou i opakovat. To se dá vyjádřit jako počet variací druhé třídy z devíti prvků s opakováním:

$V_2^*(9) = 9^2 = 81$

Takže výsledný počet pro 2 trojky: $x_2 = P^*(4).V_2^*(9) = 6.81 = 486$

Tzn., že počet státních poznávacích značek OSB XX-XX s aspoň dvěmi trojkami je:

$x = x_4 + x_3 + x_2 = 1 + 36 + 486 = 523$

Úlohy k samostatnému řešení

1.1. Zjednodušte a vypočtěte:

$$\binom{4}{2}+\binom{6}{2}-\binom{7}{2}=$$

$$\binom{6}{3}+\binom{6}{4}+\binom{7}{5}=$$

$$\frac{(n+3)!}{(n+1)!}+\frac{(n+1)!}{(n-1)!}-\frac{2(n+2)!}{n!}=$$

$$\frac{1}{n!}-\frac{3}{(n+1)!}-\frac{n^2-4}{(n+2)!}=$$

$$\frac{(n+2)!}{n!}-\frac{2(n+1)!}{(n-1)!}+\frac{n!}{(n-2)!}=$$

$$\binom{x+2}{x}+\binom{x+3}{x+1}=64$$

$$\binom{x+3}{x+1}-2\binom{x+2}{x}+3\binom{x+4}{x+2}=75$$

1.2. Kolik třítónových akordů je možné zahrát z 8 tónů?

1.3. Kolik různých optických signálů je možno dát vytahováním 5 různých barevných vlajek, je-li vždy všech pět vlajek nahoře?

1.4. Zjistěte, kolik existuje různých kvádrů, pro něž platí, že délka každé jejich hrany je přirozené číslo z intervalu $\langle 2,15\rangle$.

1.5. V obchodě mají tři druhy bonbónů v sáčcích po 100g. Kolika způsoby může zákazník koupit 1 kg bonbónů?

1.6. Kolik různých státních poznávacích značek z jedné série existuje s aspoň dvěma trojkami?

1.7. Ze 7 prvků bylo vytvořeno 2401 variací s opakováním stejné třídy. Kolik prvků obsahuje jedna variace?

1.8. Jsou dány cifry: 1, 2, 3, 4, 5. Cifry nelze opakovat. Kolik je možno vytvořit z těchto cifer čísel, která jsou

a) pětimístná, sudá

b) pětimístná, končící dvojčíslím 21

c) pětimístná, menší než 30 000

d) trojmístná, lichá

e) čtyřmístná, větší než 2000

f) čtyřmístná, začínající cifrou 2

g) čtyřmístná, sudá nebo končící cifrou 3

h) dvojmístná nebo trojmístná

1.9. Jsou dány cifry: 0, 1, 2, 3, 4. Splňte úkoly minulé úlohy (1.8.) tak, že cifry se nesmí opakovat a číslo nemůže začínat nulou.

1.10. Kolik prvků obsahuje množina všech pěticiferných přirozených čísel?

1.11. Kolik různých značek teoreticky existuje v Morseově abecedě, sestavují-li se tečky a čárky do skupin po jedné až pěti?

1.12. Kolik prvků dá 120 kombinací druhé třídy s opakováním?

1.13. Kolik je dáno prvků, jestliže variací třetí třídy z nich utvořených je pětkrát více než variací druhé třídy?

1.14. Z kolika prvků lze vytvořit 90 variací druhé třídy?

1.15. Z kolika prvků lze vytvořit 55 kombinací druhé třídy?

1.16. Zmenší-li se počet prvků o dva, zmenší se počet permutací čtyřicetdvakrát. Určete počet prvků.

1.17. Z kolika prvků lze vytvořit padesátkrát více variací třetí třídy než variací druhé třídy?

1.18. Zvětší-li se počet prvků o dva, zvětší se počet kombinací druhé třídy o 17. Určete počet prvků.

1.19. Zvětší-li se počet prvků o 8, zvětší se počet kombinací druhé třídy jedenáctkrát. Určete počet prvků.

1.20. Zmenší-li se počet prvků o 1, zmenší se počet permutací z těchto prvků desetkrát. Určete počet prvků.

1.21. Kolik permutací z n prvků $a_1, a_2, \ldots, a_n$ obsahuje prvek a_1 na prvé pozici.?

1.22. V prodejně si můžete vybrat ze sedmi druhů pohlednic. Kolika způsoby lze koupit

a) 10 pohlednic,

b) 5 pohlednic,

c) 5 různých pohlednic?

1.23. V knihkupectví prodávají 10 titulů knižních novinek. Kolika způsoby lze koupit

a) 4 knižní novinky,

b) 5 různých knižních novinek?

1.24. Na hokejovém turnaji, kterého se účastní 8 družstev, sehraje každý tým s ostatními právě 1 utkání. Kolik zápasů bude celkem sehráno?

1.25. Z 5 bílých a 4 červených kuliček tvoříme trojice tak, aby v každé trojici byly vždy 2 bílé a 1 červená kulička.. Kolik trojic splňujících tuto podmínku lze vytvořit?

1.26. Hokejový tým odjel na OH s 23 hráči, a to s 12 útočníky, 8 obránci a 3 brankáři. Kolik různých sestav může trenér teoreticky vytvořit?

1.27. Kolika přímkami lze spojit 7 bodů v rovině, jestliže

a) žádné tři z nich neleží v přímce,

b) tři z nich leží v jedné přímce?

1.28. Kolik kružnic je určeno 10 body v rovině, jestliže žádné tři z nich neleží na přímce a žádné čtyři z nich neleží na kružnici?

1.29 Kolik různých hodů můžeme provést

a) dvěma,

b) třemi různobarevnými kostkami?

1.30. V turistickém oddílu "Hbitý svišť" je 10 dívek a 8 chlapců. Určete, kolika způsoby mohou sestavit volejbalový tým (má šest členů), ve kterém budou hrát

a) právě dvě dívky.

b) maximálně dva chlapci?

1.31. Kolik prvků obsahuje množina všech pěticiferných přirozených čísel?

1.32. Deset přátel si vzájemně poslalo pohlednice z prázdnin. Kolik pohlednic celkem rozeslali?

1.33. Kolikrát více je variací k-té třídy z n prvků než kombinací k-té třídy z těchto prvků?

1.34. V plně obsazené lavici sedí 6 žáků a, b, c, d, e, f.

a) Kolika způsoby je lze přesadit?

b) Kolika způsoby je lze přesadit tak, aby žáci a, b seděli vedle sebe?

c) Kolika způsoby je lze přesadit tak, aby žák c seděl na kraji?

d) Kolika způsoby je lze přesadit tak, aby žák c seděl na kraji a žáci a, b seděli vedle sebe?

1.35. Student má v knihovně 4 různé učebnice pružnosti, 3 různé učebnice matematiky a 2 různé učebnice angličtiny. Kolika způsoby je lze seřadit, mají-li zůstat učebnice jednotlivých oborů vedle sebe?

1.36. Kolika způsoby lze rozdělit 8 účastníků finále v běhu na 100 m do 8 drah?

1.37. Kolik různých permutací lze vytvořit použitím všech písmen slova

a) statistika,

b) matematika?

1.38. Kolik různých signálů je možno vytvořit použitím pěti různobarevných praporků, použijeme-li

a) pouze 3 praporky,

b) 2 praporky?

1.39. Četa vojáků má vyslat na stráž 4 muže. Kolik mužů má četa, je-li možno úkol splnit 210 způsoby?

1.40. Kolik úhlopříček má konvexní n-úhelník?

1.41. V zásobníku je 7 ostrých a 3 slepé náboje. Určete, kolika způsoby lze namátkou ze zásobníku vyjmout 5 nábojů, z nichž alespoň 3 jsou ostré.

1.42. Kolika způsoby je možno na čtvercové šachovnici s 64 poli vybrat 3 pole tak, aby všechna tři pole neměla stejnou barvu?

1.43. Kolika způsoby je možno na šachovnici s 64 poli vybrat 3 pole tak, aby všechna neležela v jednom sloupci?

1.44. V prostoru jsou dány 2 mimoběžky a, b. Na přímce a je dáno m různých bodů $A_1, \dots A_m$, na přímce b n různých bodů $B_1, \dots, B_n$. Určete počet všech čtyřstěnů, jejichž všechny vrcholy leží na přímkách a, b, a to v bodech A_i, B_j.

Výsledky úloh k samostatnému řešení

1.1. 0, 56, 2, 0, 2, 6, 4

1.2. 56

1.3. 120

1.4. 560

1.5. 66

1.6. 523

1.7. 4

1.8. 48, 6, 48, 36, 96, 24, 72, 80

1.9. 60, 4, 48, 18, 72, 24, 78, 64

1.10. 90 000

1.11. 62

1.12. 15

1.13. 7

1.14. 10

1.15. 11

1.16. 7

1.17. 52

1.18. 8

1.19. 4

1.20. 10

1.21. (n-1)!

1.22. $C_{10}(16)$; $C_5(11)$; 21

1.23. $C_4(13)$; $C_5(10)$

1.24. 28

1.25. 40

1.26. 18 480

1.27. 21; 19

1.28. 120

1.29. 36; 216

1.30. 3150; 8106

1.31. 90 000

1.32. 90

1.33. k!

1.34. 720; 240; 240; 96

1.35. 1 728

1.36. 40 320

1.37. 75 600 , 151200

1.38. 60; 20

1.39. 10

1.40. n/2*(n-3)

1.41. 231

1.42. 31 744

1.43. 41 216

1.44. $C_2(m).C_2(n)$

2. PRAVDĚPODOBNOST JEVŮ

Průvodce studiem

V první kapitole jste se seznámili s kombinatorikou. Tyto znalosti použijeme v této kapitole, zavedeme pojem pravděpodobnost jevů a ukážeme základní metody výpočtu pravděpodobnosti.

Předpokládané znalosti

Množiny, množinové operace, pojmy z kombinatoriky.

Cíle

Cílem této kapitoly je objasnit pojmy náhodný pokus, náhodný jev, zavést operace s jevy a zformulovat základní definice pravděpodobnosti.

Výklad

2.1. Náhodný pokus, náhodný jev

Teorie pravděpodobnosti vychází ze studia náhodných pokusů.

Náhodný pokus

- je proces, který při opakování dává ze stejných podmínek rozdílné výsledky.

Výsledek pokusu není předem znám (výsledek není jednoznačně určen jeho podmínkami), je to však právě jeden z prvků známé množiny výsledků, kterou nazýváme **základní prostor Ω**

Prvky základního prostoru

(tj. možné výsledky náhodného pokusu) se nazývají **elementární náhodné jevy** (E_1, E_2, ..., E_n)

Tedy: každá podmnožina základního prostoru Ω se nazývá **náhodný jev** (značíme A, B, ...), přičemž prázdná podmnožina se nazývá **jev nemožný**, označujeme $\emptyset$ a celý základní prostor **jev jistý**, označujeme I.

Řešené úlohy

Příklad 2.1.1. Klasickým příkladem náhodného pokusu je hod hrací kostkou, tedy:

Řešení:

Náhodný pokus . . . hod hrací kostkou

Elementární jevy . . . "padne 1" ... E_1

"padne 2" ... E_2

. . .

"padne 6" ... E_6

Jevy E_1, E_2, ..., E_6 vymezují základní prostor Ω.

V tomto základním prostoru mohou být například následující jevy:

náhodný jev A . . . "padne liché číslo" . . . $A = E_1 + E_3 + E_5$

náhodný jev B . . . "padne číslo ≥ 4" . . . $A = E_4 + E_5 + E_6$

jev nemožný . . . . ."padne číslo > 6"

jev jistý . . . ."padne číslo < 7"

neslučitelné jevy. . ."padne sudé číslo", "padne liché číslo"

2.1.1. Operace s jevy

- **Součet jevů *A*, *B***

jev, který nastane právě tehdy, když nastane alespoň jeden z jevů A, B. Zavádíme označení $A+B$ nebo množinově $A \cup B$.

- **Součin jevů *A*, *B***

jev, který nastane právě tehdy, když nastanou oba jevy současně. Zavádíme označení $A.B$ nebo množinově $A \cap B$.

- **Rozdíl jevů *A*, *B***

jev, který nastane právě tehdy, když nastane jev A a nenastane jev B. Zavádíme označení $A - B$.

- Jev $\overline{A}$ nazýváme **jevem opačným** k jevu A, je-li $\overline{A} = \Omega - A$.

- Náhodné jevy se nazývají **neslučitelné** (disjunktní), jestliže platí $A.B = \emptyset$.

- Jevy $A_1, A_2, ..., A_n$ tvoří **systém neslučitelných jevů**, je-li $A_i \,.\, A_j = 0$ pro všechna $i \neq j$.
- Tento systém se nazývá **úplný**, je-li $A_1 + A_2 + ... + A_n = I = \Omega$.

2.2. Axiomatické zavedení pravděpodobnosti

Axiomatická výstavba teorie pravděpodobnosti, která pochází od významného ruského matematika A. N. Kolmogorova, vychází z toho, že pravděpodobnost je objektivní vlastnost náhodného jevu, která nezávisí na tom, zda ji umíme nebo neumíme měřit.

Definice 2.2.1.

Jevové pole $\mathcal{A}$ je množina všech různých podmnožin základního prostoru Ω, která vyhovuje těmto podmínkám:

- I leží v $\mathcal{A}$

- Leží-li jevy A, B v $\mathcal{A}$, pak $A+B$, $A.B$ i $\overline{A}$, $\overline{B}$ leží v $\mathcal{A}$

Poznámka

Na jevové pole $\mathcal{A}$ *se můžeme dívat jako na množinu jevů, ve které každý výsledek definovaných operací náleží opět do této množiny.*

Definice 2.2.2.

Nechť $\mathcal{A}$ je jevové pole. **Pravděpodobnost jevu** $\boldsymbol{A}$ je reálné číslo $P(A)$, pro něž platí:

1. $P(A) \geq 0$ . . . axiom nezápornosti
2. $P(I) = 1$ . . . axiom jednotky
3. $P(A_1 + A_2 + ... + A_n + ...) = P(A_1) + P(A_2) + ...P(A_n) + ...$, přičemž $A_1, A_2, ..., A_n, ... \in \mathcal{A}$ tvoří skupinu navzájem neslučitelných jevů . . . axiom aditivity

Věta 2.2.1. o vlastnostech pravděpodobnosti

1. $P(\emptyset) = 0$
2. $P(\overline{A}) = 1 - P(A)$
3. Jestliže $A \subseteq B$, pak:

   a) $0 \leq P(A) \leq P(B)$

   b) $P(B - A) = P(B) - P(A)$

4. $P(A + B) = P(A) + P(B) - P(A.B)$

Důkaz:

ad 1. Jev nemožný $\emptyset$ a jev jistý I jsou neslučitelné jevy. Platí: $\emptyset + I = I$ a z axiomu aditivity plyne, že

$P(I) = P(\emptyset + I) = P(\emptyset) + P(I)$ a odtud $P(\emptyset) = P(I) - P(I) = 0$

ad 2. A, $\overline{A}$ jsou neslučitelné jevy. Zároveň platí $A + \overline{A} = I$. Z axiomů jednotky a aditivity plyne:

$P(I) = P(A + \overline{A}) = 1$, takže $P(\overline{A}) = 1 - P(A)$

ad 3. Nechť $A \subset B$. Jelikož A, $\overline{A}$ jsou neslučitelné jevy, jsou neslučitelné také jevy $A.B$, $\overline{A}.B$, neboť platí

$(A.B).(\overline{A}.B) = (B.A).(\overline{A}.B) = B(A.\overline{A}).B = B.\emptyset.B = 0.$

Jev B můžeme zapsat ve tvaru $B = I.B = (A + \overline{A}).B = A.B + \overline{A}.B = A + \overline{A}.B$, neboť podle předpokladu $A \subset B$. Tedy:

$P(B) = P(A + \overline{A}.B) = P(A) + P(\overline{A}.B) \geq P(A) \geq 0.$

Protože $\overline{A}.B = B - A$, platí $P(B - A) = P(B) - P(A)$.

ad 4. Platí, že:

$A = A.I = A.(B+\overline{B}) = A.B+A.\overline{B}$

$B = B.I = B.(A+\overline{A}) = B.A+B.\overline{A}$, tudíž

$A+B = A.B+A.\overline{B} + \overline{A}.B$

Jelikož jsou jevy $A.B$, $A.\overline{B}$, $\overline{A}.B$ vzájemně neslučitelné, z axiomu aditivity vyplývá:

$P(A) = P(A.B+A.\overline{B}) = P(A.B) + P(A.\overline{B})$.

Vyjádříme-li nyní z předchozí rovnice $P(A.\overline{B})$, obdržíme:

$P(A.\overline{B}) = P(A)-P(A.B)$, obdobně:

$P(B) = P(A.B+\overline{A}.B) = P(A.B) + P(\overline{A}.B)$, tedy

$P(\overline{A}.B) = P(B)-P(A.B)$, tzn.

$P(A+B) = P(A.B+A.\overline{B}+\overline{A}.B) = P(A.B) + P(A.\overline{B}) + P(\overline{A}.B) =$

$$= P(A.B) + P(A) - P(A.B) + P(B) - P(A.B) = P(A) + P(B) - P(A.B).$$

Jsou-li jevy A, B neslučitelné, pak $A.B = \emptyset$ a uvedený vztah odpovídá axiomu aditivity.

2.3. Klasická definice pravděpodobnosti

Definice 2.3.1.

Nechť je dáno n elementárních jevů $E_1, E_2, \ldots, E_n$, které tvoří úplný systém neslučitelných jevů a jsou **stejně možné**. Rozkládá-li se jev A na m ($m \leq n$) elementárních jevů z tohoto systému, pak pravděpodobnost jevu A je reálné číslo $P(A) = \frac{m}{n}$

Poznámka

Klasická definice pravděpodobnosti se užívá, je-li:

konečný počet elementárních jevů

stejná míra výskytu elementárních jevů

Všechny elementární jevy se obvykle označují jako všechny možné případy. Všechny elementární jevy, na které se rozkládá jev A, se nazývají všechny příznivé případy. Pak daný vztah přejde na známý tvar:

$$P(A) = \frac{\text{počet všech příznivých případů}}{\text{počet všech možných případů}}$$

Řešené úlohy

Příklad 2.3.1. Rozhodněte, zda v následujících případech je stejná míra výskytu elementárních jevů:

a) hod navrtanou kostkou

b) hod mincí

c) výstřel do terče

Řešení:

ad *a*) E_1 - padne 1, E_2 - padne 2, ..., E_6 - padne 6, není stejná míra výskytu

ad *b*) E_1 - padne rub, E_2 - padne líc, je stejná míra výskytu

ad *c*) E_1 - zásah, E_2 - mimo, u většiny střelců není stejná míra výskytu

Příklad 2.3.2. Při hodu kostkou určete pravděpodobnost jevů:

a) jev *A*: "padne číslo 5"

b) jev *B*: "padne číslo ≤ 2"

Řešení:

ad *a*) $P(A) = \frac{1}{6}$

ad *b*) $P(B) = \frac{2}{6} = \frac{1}{3}$

Příklad 2.3.3. S jakou pravděpodobností padne na dvou kostkách součet

a) šest

b) menší než 7

Řešení:

ad *a*) Šestka padne v následujících případech:

1. kostka	1	5	2	4	3
2. kostka	5	1	4	2	3

Tzn. 5 možností, $m = 5$

Počet všech možností: $n = \binom{6}{1} . \binom{6}{1} = 36$

$$P(A) = \frac{m}{n} = \frac{5}{36} = 0{,}13\overline{8}$$

ad *b*)

Z předchozího vyplývá, že je 5 možností pro součet šest. Ostatní možnosti:

	součet 5				**součet 4**			**součet 3**		**součet 2**
1. kostka	1	4	2	3	1	3	2	1	2	1
2. kostka	4	1	3	2	3	1	2	2	1	1

Takže $m = 5 + 4 + 3 + 2 + 1 = 15$

$$P(B) = \frac{m}{n} = \frac{15}{36} = 0{,}41\overline{6}$$

Příklad 2.3.4. V cele předběžného zadržení sedí vedle sebe 10 podezřelých, z toho 3 ženy. Jaká je pravděpodobnost, že všechny tři ženy sedí vedle sebe?

Řešení: Počet možností, jak uspořádat 10 podezřelých, odpovídá počtu permutací z 10 prvků: $n = 10!$

$m = 8.3!.7!$ - existuje 8 způsobů umístění dané trojice žen (na pozicích 123, 234, 345, ..., 8910), 3! způsobů jak danou trojici uspořádat a 7! způsobů, jak uspořádat zbývající delikventy.

$$P(A) = \frac{8.3!.7!}{10!} = 0{,}0\overline{6}$$

Příklad 2.3.5. Stanovte pravděpodobnost jevu, že z 10 náhodně vytažených bridžových karet budou alespoň 3 esa. (bridžové karty: 52 karet celkem, z toho 4 esa)

Řešení: Jev A - vybereme alespoň 3 esa, znamená, že vybereme 3 nebo 4 esa. To znamená, že jev A se rozkládá na součet dvou navzájem disjunktních jevů:

A_1 . . . vybereme 3 esa

A_2 . . . vybereme 4 esa

$P(A) = P(A_1 + A_2) = P(A_1) + P(A_2)$, kde:

$$P(A_1) = \frac{m_1}{n} = \frac{C_3(4).C_7(48)}{C_4(52)} = \frac{\binom{4}{3}.\binom{48}{7}}{\binom{52}{10}}$$

Hodnotu n (počet všech možných případů) jsme vypočetli pomocí kombinací bez opakování - z 52 karet vybíráme čtyři bez ohledu na pořadí, přičemž karty nevracíme zpět.

Hodnotu m_1 (počet všech příznivých případů) jsme vypočetli podobnou úvahou: ze čtyř es vybíráme tři bez ohledu na pořadí a ze zbývajících 48 karet vybíráme sedm, opět bez zřetele na uspořádání.

Zcela analogicky vypočteme

$$P(A_2) = \frac{m_2}{n} = \frac{C_4(4).C_6(48)}{C_4(52)} = \frac{\binom{4}{4}.\binom{48}{6}}{\binom{52}{10}}$$

Takže:

$$P(A) = \frac{m_1 + m_2}{n} = \frac{\binom{4}{3}.\binom{48}{7} + \binom{4}{4}.\binom{48}{6}}{\binom{52}{10}} = 0,019$$

Příklad 2.3.6. Při slosování sportky je z osudí postupně vylosováno 6 čísel ze 49. Po vylosování těchto čísel je ze zbývajících čtyřiceti tří čísel vylosováno dodatkové číslo. Při správném tipování:

a) šesti čísel, získává sázející výhru 1. pořadí,

b) pěti čísel a dodatkového čísla (5 + 1), získává sázející výhru 2. pořadí,

c) pěti čísel, získává sázející výhru 3. pořadí,

d) čtyř čísel, získává sázející výhru 4. pořadí,

e) tří čísel, získává sázející výhru 5. pořadí.

Vypočtěte pravděpodobnost, se kterou při vsazeném jednom sloupci vyhrajete v 1.tahu výhry *a* - *e*.

Řešení: Řešit budeme obdobně, jako předchozí příklad 2.3.5.

ad *a*)

$$P(A_1) = \frac{\binom{6}{6}.\binom{43}{0}}{\binom{49}{6}} = \frac{1}{13983816} = 7,15.10^{-8}$$

(řádově se jedná o stejnou pravděpodobnost, s jakou v ruletě padne pětkrát po sobě stejné číslo: $(1/37)^5 = 1,44.10^{-8}$)

ad b)

$$P\left(A_2\right)=\frac{\binom{6}{5}\cdot\binom{1}{1}\cdot\binom{42}{0}}{\binom{49}{6}}=\frac{6}{13983816}=4,2.10^{-7}$$

ad c)

$$P\left(A_3\right)=\frac{\binom{6}{5}\cdot\binom{43}{1}\cdot\binom{1}{0}}{\binom{49}{6}}=\frac{252}{13983816}=1,802.10^{-5}$$

ad d)

$$P\left(A_4\right)=\frac{\binom{6}{4}\cdot\binom{43}{2}}{\binom{49}{6}}=\frac{13545}{13983816}=0,000969$$

ad e)

$$P\left(A_5\right)=\frac{\binom{6}{3}\cdot\binom{43}{3}}{\binom{49}{6}}=\frac{246820}{13983816}=0,0177$$

2.4. Geometrická pravděpodobnost

Geometrická pravděpodobnost

- používáme ji v případech, které lze převést na toto schéma:

V rovině (případně na přímce nebo v prostoru) je dána určitá oblast Ω a v ní další uzavřená oblast A.

Pravděpodobnost jevu A, který spočívá v tom, že náhodně zvolený bod v oblasti Ω leží i v oblasti A je:

$P\left(A\right)=\frac{|A|}{|\Omega|}$, kde $|A|$, $|\Omega|$ jsou míry oblastí A a Ω

Řešené úlohy

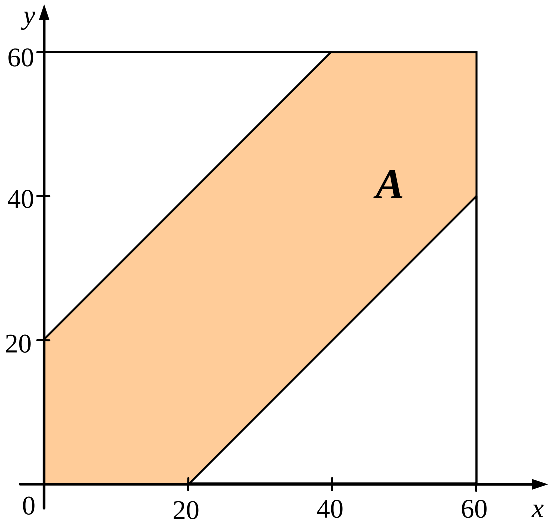
Příklad 2.4.1. Jak je pravděpodobné, že meteorit padne na pevninu, víme-li, že pevnina má rozlohu 149 milionů km^2 a moře 361 milionů km^2.

Řešení:

$$P(A) = \frac{149}{149+361} = 0,292$$

Příklad 2.4.2. Dva známí se domluví, že se sejdou na určitém místě mezi 15. a 16. hodinou, přičemž doba čekání je 20 minut. Jaká je pravděpodobnost, že se při této dohodě setkají?

Řešení:



x . . . doba po 15.hodině v níž přijde první,

$x \in \langle 0,60 \rangle$

y . . . doba po 15.hodině v níž přijde druhý,

$x \in \langle 0,60 \rangle$

jev A . . . oblast vymezená čtvercem a nerovnicí

$|x - y| \leq 20$

$|\Omega| = 60.60 = 3600$

Když spojíme dva nevyšrafované trojúhelníky, tak dostaneme čtverec o straně délky 40, tedy:

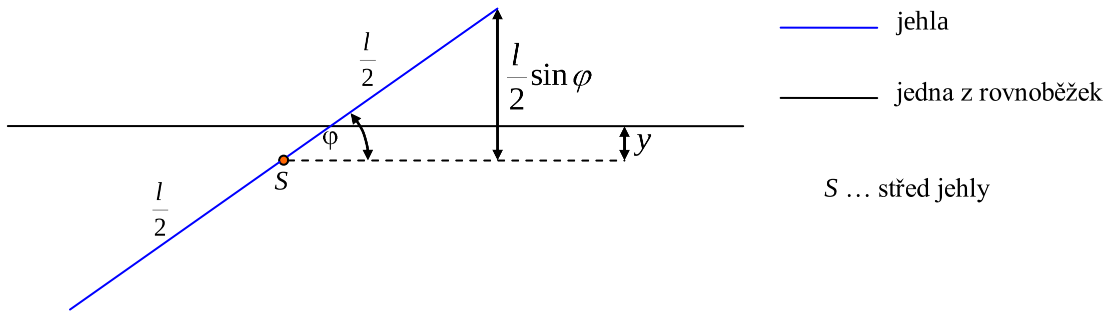
$|A| = 3600 - 40.40 = 2000$

Takže:

$$P(A) = \frac{2000}{3600} = \frac{5}{9} = 0,56$$

Příklad 2.4.3. V rovině jsou narýsovány rovnoběžky, jejichž vzdálenost je d. Určete pravděpodobnost toho, že náhodně vržená jehla délky l ($l < d$) protne libovolnou přímku.

Řešení: Situace je vystižena na obrázku:



Každou polohu jehly můžeme tedy popsat dvěmi souřadnicemi: vzdáleností y jejího středu S od nejbližší z přímek a úhlem φ jehly s daným systémem přímek.

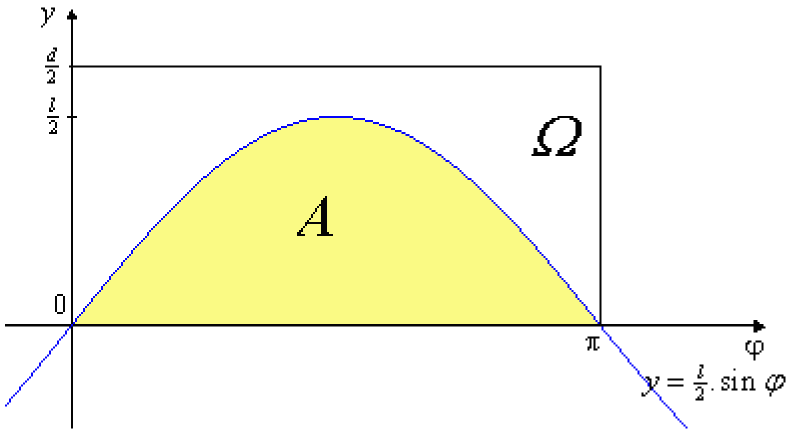
Platí: $0 \le y \le \frac{d}{2}$; $0 \le \varphi \le \pi$

Jehla protne nejblíže položenou přímku, jestliže:

$\frac{l}{2}.\sin\varphi \ge y$ (vymezení oblasti A)

Možným souřadnicím středu jehly odpovídá pravoúhelník

$\Omega = \langle 0, \pi \rangle \times \left\langle 0, \frac{d}{2} \right\rangle$ viz. obr.



Z předchozího vyplývá, že:

$$|\Omega| = \pi \frac{d}{2}$$

$$|A| = \int_0^{\pi} \frac{l}{2}.\sin\varphi \, d\varphi = \left[-\frac{l}{2}.\cos\varphi \right]_0^{\pi} = \frac{l}{2} + \frac{l}{2} = l$$

Tedy:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2l}{\pi d}$$

Tzn. jestliže např. $d = 2$, $l = 1$, pak

$$P(A) = \frac{2}{2\pi} = \frac{1}{\pi} = 0{,}318$$

2.5. Statistická definice pravděpodobnosti

Definice 2.5.1.

Nechť A je hromadný jev. Nastane-li v n pokusech jev A právě f_n krát, definujeme:

$$P(A) = \lim_{n \to \infty} \frac{f_n}{n}$$

Číslo f_n se nazývá absolutní četnost jevu A, $\frac{f_n}{n}$ - relativní četnost jevu A při n pokusech

Hromadný jev

jev, který lze za daného systému podmínek libovolně krát opakovat nebo který lze pozorovat na hromadně se vyskytujících předmětech téhož druhu

Řešené úlohy

Příklad 2.5.1. Při házení mincí byly zjištěny tyto výsledky:

Řešení:

počet hodů n	počet padnutí líce f_n	relativní četnost $\frac{f_n}{n}$
4000	2032	0,5080
12000	6019	0,5016
24000	12012	0,5005
30000	15010	0,5003

Z tabulky je zřejmé, že platí:

$$P(A) = \lim_{n\to\infty} \frac{f_n}{n} = 0{,}5$$

2.6. Podmíněná pravděpodobnost a nezávislé jevy

Definice 2.6.1.

Pravděpodobnost uskutečnění jevu *A* za předpokladu, že nastal jev *B*, se zapisuje *P*(*A*/*B*) a nazývá se podmíněná pravděpodobnost. Je rovna:

$$P(A/B) = \frac{P(A.B)}{P(B)}$$

Řešené úlohy

Příklad 2.6.1. Házíme dvěma mincemi.

Jev *A*: padne líc a rub

Jev *B*: na první minci padne líc

Určete pravděpodobnost jevu *A* za předpokladu, že nastal jev *B*.

Řešení: Možnosti, které mohou nastat:

RUB RUB

RUB LÍC

LÍC RUB

LÍC LÍC

a) pomocí klasické definice: *P*(*A* / *B*) = 0,5

b) pomocí vzorce na podmíněnou pravděpodobnost: $P(A/B) = \frac{P(A.B)}{P(B)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$

Příklad 2.6.2. Máme krabici se třemi bílými a dvěma černými koulemi. Vytáhneme postupně dvě koule (první nevracíme zpět). Určete pravděpodobnost toho, že v druhém tahu vytáhneme bílou kouli za předpokladu, že v prvním tahu byla vytažena černá koule.

Řešení:

jev *A*: ve druhém tahu vytažena bílá

jev *B*: v prvním tahu vytažena černá

Možnosti:

	1. tah	**2. tah**	**celkem**
počet možností	černá $\binom{2}{1}$	černá $\binom{1}{1}$	2
	černá $\binom{2}{1}$	bílá $\binom{3}{1}$	6
	bílá $\binom{3}{1}$	černá $\binom{2}{1}$	6
	bílá $\binom{3}{1}$	bílá $\binom{2}{1}$	6

Z tabulky vidíme, že:

$$P(A.B)=\frac{6}{20}$$

$$P(B)=\frac{8}{20}$$

To znamená: $P(A/B)=\frac{P(A.B)}{P(B)}=0,75$

Věta 2.6.1.

Pro pravděpodobnost součinu dvou jevů *A*, *B* platí:

$P(A.B) = P(A).P(B / A) = P(B).P(A / B)$

Důkaz: Tvrzení plyne přímo z definice 2.6.1.

Definice 2.6.2.

Dva jevy *A*, *B* nazýváme **nezávislé**, jestliže platí: $P(A / B)=P(A)$

Poznámky:

Jsou-li jevy A, B nezávislé, pak $P(A.B) = P(A).P(B)$.

Pojem nezávislosti není totožný s pojmem neslučitelnosti.

Jsou-li A, B neslučitelné jevy, pak $P(A+B) = P(A)+P(B)$.

U skupiny více než dvou jevů rozlišujeme nezávislost podvojnou a vzájemnou

Jevy A_1, ..., A_n jsou vzájemně nezávislé, jestliže pro každou jejich podmnožinu platí, že pravděpodobnost průniku jevů je rovna součinu pravděpodobností těchto jevů.

Jsou-li jevy vzájemně nezávislé, jsou také po dvou nezávislé. Opačné tvrzení neplatí!

Řešené úlohy

Příklad 2.6.3. Studenti při zkoušení mohou dostat tři otázky. První student je připraven pouze na první otázku, druhý umí pouze druhou otázku, třetí ovládá jen třetí otázku a čtvrtý je připraven na všechny tři otázky. Uvažujme nyní tyto jevy:

A_1 . . . vyvolaný student dokáže zodpovědět první otázku

A_2 . . . vyvolaný student dokáže zodpovědět druhou otázku

A_3 . . . vyvolaný student dokáže zodpovědět třetí otázku

Ukažte, že jevy A_1, A_2, A_3 jsou po dvou nezávislé, ale nejsou vzájemně nezávislé.

Řešení: Z klasické definice pravděpodobnosti plyne, že:

$P(A_1) = P(A_2) = P(A_3) = 2/4 = 0{,}5.$

Uvažujme nyní jevy: $A_1.A_2$, $A_1.A_3$, $A_2.A_3$, $A_1.A_2.A_3$.

Pro pravděpodobnosti těchto jevů opět z klasické definice pravděpodobnosti vyplývá:

$P(A_1.A_2) = P(A_1.A_3) = P(A_2.A_3) = P(A_1.A_2.A_3) = 0{,}25.$

Pro jednotlivé dvojice jevů tedy platí:

$P(A_i.A_j) = P(A_i).P(A_j) = 0{,}5.0{,}5 = 0{,}25 \ (i \neq j)$

Takže jevy A_1, A_2, A_3 jsou po dvou nezávislé.

Vzhledem k tomu, že $P(A_1.A_2.A_3) \neq P(A_1).P(A_2).P(A_3)$, neboť $0{,}25 \neq 0{,}5.0{,}5.0{,}5$, nejsou tyto tři jevy vzájemně nezávislé.

2.7. Úplná pravděpodobnost a Bayesova věta

Řešené úlohy

Příklad 2.7.1. V obchodě jsou tři pokladny na nichž dojde k chybě v účtování s pravděpodobností: 0,1; 0,05 a 0,2, přičemž z hlediska umístění pokladen v obchodě jsou pravděpodobnosti odbavení pokladnami 0,3; 0,25 a 0,45. Jaká je pravděpodobnost, že osoba opouštějící obchod má chybný účet?

Řešení:

jev A: došlo k chybě v účtování

jev H_i: odbavení i-tou pokladnou

jev A je možno vyjádřit:

$A = A.H_1 + A.H_2 + A.H_3$

(zákazník má chybný účet, přičemž projde první pokladnou nebo má chybný účet po odbavení druhou pokladnou nebo má chybný účet a prošel třetí pokladnou)

Jevy $A.H_1$, $A.H_2$, $A.H_3$ jsou vzájemně neslučitelné, proto:

$P(A) = P(A.H_1 + A.H_2 + A.H_3) = P(A.H_1) + P(A.H_2) + P(A.H_3) =$ (z věty 2.6.1.)

$= P(H_1).P(A/H_1) + P(H_2).P(A/H_2) + P(H_3).P(A/H_3) =$

$= 0{,}3.0{,}1 + 0{,}25.0{,}05 + 0{,}45.0{,}2 = 0{,}1325$

Zobecněním postupu z předchozí úlohy řešíme úlohy formulované na základě **výchozí situace**:

- Máme určit pravděpodobnost jevu A, o kterém je známo, že může nastat pouze současně s některým z jevů $H_1, H_2, ..., H_n$, které tvoří úplný systém neslučitelných jevů:

Věta 2.7.1. (o úplné pravděpodobnosti)

Nechť je dán úplný systém vzájemně neslučitelných jevů $H_1, H_2, ..., H_n$ a libovolný jev A, který může nastat pouze současně s některým z jevů H_i. Pro pravděpodobnost jevu A platí:

$$P(A) = P(H_1).P(A/H_1)+P(H_2).P(A/H_2)+...+P(H_n).P(A/H_n) = \sum_{i=1}^{n} P(H_i).P(A/H_i)$$

Důkaz: Zjevný, zobecněním postupu v příkladu 2.7.1. na n jevů $H_1, H_2, ..., H_n$

Řešené úlohy

Příklad 2.7.2. Zadání je stejné jako v předchozím příkladě. Otázka: Jaká je pravděpodobnost, že jsme byli u druhé pokladny, máme-li chybný účet?

Řešení: Hledáme tedy, čemu je rovno $P(H_2 / A)$. Lehce odvodíme:

$$P(H_2/A) = \frac{P(H_2.A)}{P(A)} = \frac{P(H_2).P(A/H_2)}{P(A)} = \frac{0{,}25.0{,}05}{0{,}1325} = 0{,}094$$

Tato situace se dá opět shrnout:

Věta 2.7.2. - Bayesova věta

Nechť je dán úplný systém vzájemně neslučitelných jevů H_1, H_2, ..., H_n a libovolný jev A, který může nastat jen současně s některým z jevů H_i. Pak pravděpodobnost, že nastane jev H_i, za předpokladu, že nastal jev A je:

$$P(H_i/A) = \frac{P(H_i).P(A/H_i)}{P(A)}, \text{ kde } P(A) = \sum_{k=1}^{n} P(H_k).P(A/H_k)$$

Důkaz: Opět zjevné, viz. předchozí příklad 2.7.2.

2.8. Opakované pokusy

Stává se, že náhodný pokus, jehož výsledkem je jev A, opakujeme n-krát po sobě při zachování stejného systému podmínek. Pokud pravděpodobnost jevu A při každém opakování nezávisí na výsledcích předcházejících pokusů, hovoříme o **Bernoulliho posloupnosti nezávislých pokusů** (např. hod kostkou). **Závislými** pak nazveme takové opakované pokusy, při nichž je pravděpodobnost "nastoupení" jevu A v určitém pokusu závislá na výsledcích předchozích pokusů (např. výběry z osudí bez vracení).

2.8.1. Nezávislé pokusy

Řešené úlohy

Příklad 2.8.1. Házíme šestkrát kostkou. Vypočtěte pravděpodobnost, že z těchto šesti hodů padne šestka právě dvakrát.

Řešení: Jedna z možností, které mohou nastat je, že šestka padne na první a druhé kostce, přičemž na zbývajících kostkách padne jakékoliv číslo vyjma šestky: 66XXXX. Pravděpodobnost, že tato situace nastane, se vypočte jakou součin pravděpodobností, s jakou padnou čísla na jednotlivých kostkách:

$$\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}=\left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{6}\right)^4$$

Další možnosti, kdy padnou dvě šestky jsou stejně pravděpodobné jako první možnost. Jedná se o případy:

66XXXX

6X6XXX

.

.

.

XXX6X6

XXXX66

... počet všech těchto možností lze vypočíst např. pomocí permutací s opakováním:

$$P^*(6)=\frac{6!}{2!.4!}=\frac{6!}{2!.(6-2)!}=\binom{6}{2}$$

Hledaná pravděpodobnost je tedy dána vztahem:

$$P=\binom{6}{2}.\left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{6}\right)^4$$

Pokud naše úvahy z předchozího příkladu shrneme, obdržíme:

Věta 2.8.1.

Je-li pravděpodobnost jevu A v každém pokusu $P(A) = p$, pak pravděpodobnost jevu A_k, že se jev A v Bernoulliho posloupnosti n nezávislých pokusů uskuteční právě k-krát, je určena vztahem:

$$P(A_k)=\binom{n}{k}.p^k.(1-p)^{n-k}$$

Důkaz: Vyjdeme z řešení příkladu 2.8.1.. Výraz p^k vyjadřuje pravděpodobnost, že jev A nastal právě v k pokusech. Výraz $(1 - p)^{n-k}$ vyjadřuje pravděpodobnost, že jev A nenastal právě v $n - k$ pokusech. V celé posloupnosti n pokusů může jev A nastat celkem $\binom{n}{k}$ způsoby. Proto je hledaná pravděpodobnost:

$$P(A_k)=\binom{n}{k}.p^k.(1-p)^{n-k}$$

Poznámka:

Ve vzorci z předchozí věty bychom pro různé hodnoty parametru k dostávali různé výsledky.

Někdy je účelné najít způsob, kterým zjistíme, které k má největší pravděpodobnost. K tomu užíváme vztahu:

p.(n + 1) - 1 ≤ k ≤ p.(n + 1)

Řešené úlohy

Příklad 2.8.2. Pravděpodobnost, že náhodně vybraný student bude znát učivo, je 0,005. Jaká je pravděpodobnost, že mezi dvaceti vybranými studenty bude:

a) právě 5 znalých studentů

b) nejvýše 2 znalí studenti

c) alespoň jeden znalý student

d) jaký je nejpravděpodobnější počet znalých studentů

ad *a*)

$$P(A_5)=\binom{20}{5}.0,005^5.0,995^{15}$$

ad *b*)

$$P=P(A_0)+P(A_1)+P(A_2)=$$

$$=\binom{20}{0}.0,005^0.0,995^{20}+\binom{20}{1}.0,005^1.0,995^{19}+\binom{20}{2}.0,005^2.0,995^{18}$$

ad *c*)

$$P=P(A_1)+P(A_2)+\ldots+P(A_{20})=1-P(A_0)=1-\binom{20}{0}.0,005^0.0,995^{20}$$

ad *d*)

$$p.(n+1)-1\le k\le p.(n+1)$$

$$0,005.21-1\le k\le 0,005.21$$

$$-0,895\le k\le 0,105$$

Takže nejpravděpodobnější počet znalých studentů je $k = 0$

2.8.2. Závislé pokusy

Řešené úlohy

Příklad 2.8.3. V osudí jsou 2 bílé a 3 černé koule. Vypočtěte pravděpodobnost toho, že:

a) vytáhneme 3 koule a budou 2 černé a 1 bílá

b) vytáhneme bez vracení jako první černou kouli, pak bílou a nakonec černou.

Řešení:

ad *a*) $P = \dfrac{\binom{3}{2}.\binom{2}{1}}{\binom{5}{3}} = \dfrac{3}{5}$

ad *b*) ČBČ . . . $P = \dfrac{\binom{3}{1}}{\binom{5}{1}}.\dfrac{\binom{2}{1}}{\binom{4}{1}}.\dfrac{\binom{2}{1}}{\binom{3}{1}} = \dfrac{3.2.2}{5.4.3} = \dfrac{1}{5}$

(další možná pořadí: ČČB, BČČ - obě se stejnou pravděpodobností jako ČBČ, všechny dohromady tedy dávají případ ad *a*)

Situaci z předchozího příkladu 2.8.3*a*. opět shrneme ve větě:

Věta 2.8.2.

Nechť je dán soubor N prvků, z nichž M má určitou vlastnost a (N - M) nikoliv. Vybereme postupně n prvků, z nichž **žádný nevracíme**. Pravděpodobnost, že mezi n vybranými bude k takových, že mají sledovanou vlastnost, vypočteme podle vzorce:

$$P = \frac{\binom{M}{k}.\binom{N-M}{n-k}}{\binom{N}{n}}$$

Důkaz: Zřejmé - odvozeno z klasické definice pravděpodobnosti

Řešené úlohy

Příklad 2.8.4. Mezi 15 výrobky je 5 zmetků. Vybereme 3 výrobky. Jaká je pravděpodobnost, že jeden z nich je vadný, jestliže:

a) vybereme všechny 3 najednou

b) vybíráme po jednom bez vracení

Řešení:

ad *a*) $P = \dfrac{\binom{5}{1}\cdot\binom{10}{2}}{\binom{15}{3}} = \dfrac{45}{91}$

ad *b*) Možnosti: (V-vadný, D-dobrý)

$\text{VDD} \ldots\ P_1 = \dfrac{5}{15}\cdot\dfrac{10}{14}\cdot\dfrac{9}{13} = \dfrac{15}{91}$

$\text{DVD} \ldots\ P_2 = \dfrac{10}{15}\cdot\dfrac{5}{14}\cdot\dfrac{9}{13} = \dfrac{15}{91}$

$\text{DDV} \ldots\ P_3 = \dfrac{10}{15}\cdot\dfrac{9}{14}\cdot\dfrac{5}{13} = \dfrac{15}{91}$

To jsou všechny možné způsoby výběru:

$P = P_1 + P_2 + P_3 = \dfrac{45}{91}$

Poznámka

Nezáleží tedy na tom, vybereme-li výrobky najednou nebo postupně bez vracení.

2.9. Řešené úlohy - pravděpodobnost (souhrnně)

Příklad 2.9.1. Mějme pět vstupenek po 100 Kč, tři vstupenky po 300 Kč a dvě vstupenky po 500 Kč. Vyberme náhodně tři vstupenky. Určete pravděpodobnost toho, že:

a) alespoň dvě z těchto vstupenek mají stejnou hodnotu

b) všechny tři vstupenky stojí dohromady 700 Kč

Řešení:

ad *a*)

Budeme řešit pomocí opačného jevu. Opačný jev k "alespoň dvě mají stejnou hodnotu" je "každá má jinou hodnotu":

$$P(A)=1-\frac{\binom{5}{1}\cdot\binom{3}{1}\cdot\binom{2}{1}}{\binom{10}{3}}=0,75$$

ad *b*)

Dohromady za 700 Kč, tzn. jedna za 100 Kč a dvě za 300 Kč nebo dvě za 100 Kč a jedna za 500 Kč:

$$P(B)=\frac{\binom{5}{1}\cdot\binom{3}{2}+\binom{5}{2}\cdot\binom{2}{1}}{\binom{10}{3}}=\frac{7}{24}=0,291\overline{6}$$

Příklad 2.9.2. Z celkové produkce závodu jsou 4% zmetků a z dobrých je 75% standardních. Určete pravděpodobnost, že náhodně vybraný výrobek je standardní.

Řešení:

jev *A*...vybraný výrobek není zmetek

jev *B* ...vybraný výrobek je standardní

Víme, že: $P(A) = 1 - 0,04 = 0,96$; $P(B/A) = 0,75$

Hledaná pravděpodobnost:

$P(A.B) = P(A).P(B/A) = 0,96.0,75 = 0,72$

Příklad 2.9.3. Z výrobků určitého druhu dosahuje 95% předepsanou kvalitu. V určitém závodě, který vyrábí 80% celkové produkce, však předepsanou kvalitu má 98% výrobků. Mějme náhodně vybraný výrobek předepsané kvality. Jaká je pravděpodobnost, že byl vyroben ve výše uvedeném závodě?

Řešení:

jev *A*...výrobek je vyroben ve zmiňovaném závodě

jev *B*...výrobek je předepsané kvality

$$P(A/B)=\frac{P(A.B)}{P(B)}=\frac{0,8.0,98}{0,95}=0,825$$

Příklad 2.9.4. Menza VŠB zakoupila 12 chladniček z 1. závodu, 20 z 2. závodu a 18 z 3. závodu. Pravděpodobnost, že chladnička je výborné jakosti, pochází-li z 1.závodu je

0,9, z 2.závodu 0,6 a z 3.závodu 0,9. Jaká je pravděpodobnost, že náhodně vybraná chladnička bude výborné jakosti?

Řešení:

jev A...náhodně vybraná chladnička bude výborné jakosti

jev B_i... náhodně vybraná chladnička pochází z i-tého závodu

Chladniček je dohromady 50.

$$A = (A.B_1) + (A.B_2) + (A.B_3)$$

$$P(A) = P(A.B_1) + P(A.B_2) + P(A.B_3)$$

$P(A) = P(B_1).P(A/B_1) + P(B_2).P(A/B_2) + P(B_3).P(A/B_3)$

$$P(A) = \frac{12}{50}.0,9 + \frac{20}{50}.0,6 + \frac{18}{50}.0,9 = 0,78$$

Příklad 2.9.5. Ve společnosti je 45% mužů a 55% žen. Vysokých nad 190 cm je 5 % mužů a 1 % žen. Náhodně vybraná osoba je vyšší než 190 cm. Jaká je pravděpodobnost, že je to žena?

Řešení:

jev A...vybraný člověk je vyšší než 190 cm

jev B_1...vybraný člověk je muž

jev B_2...vybraný člověk je žena

$$P(A) = P(A.B_1) + P(A.B_2) = 0,45.0,05 + 0,55.0,01 = 0,028$$

$$P(B_2/A) = \frac{P(A.B_2)}{P(A)} = \frac{0,55.0,01}{0,028} = 0,196$$

Příklad 2.9.6. Sada, kterou tvoří 100 součástek, je podrobena výběrové kontrole. Sada se nepřijme, jestliže mezi pěti kontrolovanými součástkami je alespoň jedna vadná. Jaká je pravděpodobnost toho, že se sada nepřijme, jestliže obsahuje 5% vadných součástek?

Řešení: Budeme řešit pomocí opačného jevu. Ten spočívá v tom, že sada bude přijata.

Tento jev je průnikem pěti jevů:

$\overline{A} = A_1.A_2.A_3.A_4.A_5$, kde A_k znamená, že k-tá kontrolovaná součástka je kvalitní.

Pravděpodobnost jevu A_1: $P(A_1) = \frac{95}{100}$ (100 součástek z nichž je 95 kvalitních)

Když nastane jev A_1, zůstane 99 součástek, mezi nimiž je 94 kvalitních, takže:

$$P\left(A_2\right)=\frac{94}{99}$$

Pravděpodobnost zbývajících jevů odvodíme obdobným způsobem, tzn.

$$P\left(\overline{A}\right)=\frac{95}{100}\cdot\frac{94}{99}\cdot\frac{93}{98}\cdot\frac{92}{97}\cdot\frac{91}{96}=0,77$$

$P(A) = 1 - P\left(\overline{A}\right) = 1 - 0,77 = 0,23$

Příklad 2.9.7. Dva střelci vystřelí po jedné ráně. Pravděpodobnosti zásahu cíle jsou po řadě 0,5 a 0,9. Určete pravděpodobnost toho, že alespoň jeden střelec zasáhne cíl.

Řešení:

jev A: alespoň jeden zasáhne cíl

jev B: cíl zasáhne první střelec

jev C: cíl zasáhne druhý střelec

$P(A) = P(B.\overline{C} + \overline{B}.C + B.C) = P(B.\overline{C}) + P(\overline{B}.C) + P(B.C) =$

$= P(B).P(\overline{C}) + P(\overline{B}).P(C) + P(B).P(C)$

$= 0,5.0,1 + 0,5.0,9 + 0,5.0,9 = 0,95$

nebo:

$P(A) = 1 - P(\overline{B}.\overline{C}) = 1 - P(\overline{B}).P(\overline{C}) = 1 - 0,5.0,1 = 0,95$

Příklad 2.9.8. Vypočtěte, co je pravděpodobnější? Vyhrát v tenise se stejně silným soupeřem 3 zápasy ze 4 nebo 6 zápasů z osmi?

Řešení: Tenisové zápasy jsou vlastně opakované nezávislé pokusy. Hrajeme-li se stejně silným soupeřem je pravděpodobnost výhry v každém zápase p = 0,5, takže:

Pravděpodobnost, že vyhrajeme 3 zápasy ze 4:

$$P\left(A_3\right)=\binom{4}{3}.0,5^3.0,5^1=4.0,5^4=0,25$$

Pravděpodobnost, že vyhrajeme 6 zápasů z 8:

$$P\left(A_6\right)=\binom{8}{6}.0,5^6.0,5^2=28.0,5^8 \doteq 0,109$$

Pravděpodobnější je tedy zvítězit ve třech zápasech ze čtyř.

Příklad 2.9.9. Narozeninový problém I. Spočítejte pravděpodobnost, že žádní dva lidé z patnáctičlenné skupiny nemají narozeniny ve stejný den roku. Ignorujte 29.únor.

Řešení: Označme $P(n)$...pravděpodobnost, že dva lidé z n-členné skupiny nemají narozeniny ve stejný den.

$\underline{n = 2}$

První člověk má narozeniny libovolný den v roce. Pravděpodobnost, že druhý člověk nemá narozeniny tentýž den je:

$$P(2) = \frac{364}{365}$$

$\underline{n = 3}$

Navážeme-li na předchozí úvahu, pak:

$$P(3) = \frac{364}{365} . \frac{363}{365}$$

Obdobně tedy:

$$P(4) = P(3) . \frac{362}{365}$$

$$\vdots$$

$$P(n) = \frac{P(n-1) . \left[365 - (n-1)\right]}{365}$$

$$P(n) = \frac{364.363.\ldots.\left[365 - (n-1)\right]}{365^{n-1}}$$

$$P(n) = \frac{365.364.363.\ldots.\left[365 - (n-1)\right].(365-n)!}{365.365^{n-1}.(365-n)!} = \frac{365!}{365^{n}.(365-n)!}$$

Takže jsme odvodili obecný vzorec, nyní pro $n = 15$:

$$P(15) = \frac{365!}{365^{15}.350!} = \frac{365.364.\ldots.351}{365^{15}} \doteq 0{,}747$$

Příklad 2.9.10. Narozeninový problém II. *(Richard von Mises, 1939)* Kolik lidí se musí nacházet v místnosti, aby, ignorujíce 29.únor, dva z nich měli narozeniny ve stejný den roku s pravděpodobností alespoň 50%.

Řešení: Označme $\overline{P}(n)$ ...pravděpodobnost, že dva lidé z n-členné skupiny mají narozeniny ve stejný den. Využijeme řešení předchozího příkladu. Stačí si uvědomit, že: $\overline{P}(n) = 1 - P(n)$, tedy:

$$\overline{P}(n) = 1 - \frac{365!}{365^n.(365-n)!}$$

Lehce zjistíme, že $\overline{P}(n) > 0{,}5$ poprvé pro $n = 23$ ($\overline{P}(23) = 0{,}507$)

V místnosti se tedy musí nacházet alespoň 23 lidí.

Úlohy k samostatnému řešení - tématicky tříděno

Jevová algebra

2.1. Znázorněním příslušných jevů ověřte platnost následujících vztahů mezi jevy:

a) idempotence $A + A = A$ $\quad A.A = A$

b) komutace $A + B = B + A$ $\quad A.B = B.A$

c) asociace $A + (B + C) = (A + B) + C$ $\quad A.(B.C) = (A.B).C$

d) distribuce $A.(B + C) = A.B + A.C$

e) absorbce $A + A.B = A$ $\quad A.(A + B) = A$

f) $A+\overline{A}=I \quad A.\overline{A}=\varnothing \quad A + I = I$

$A+\varnothing=A \quad A.\varnothing=\varnothing \quad A.I = A$

g) reflexe $A \subset A$

h) tranzitivnost $A \subset B, B \subset C \Rightarrow A \subset C$

i) antisymetrie $A \subset B, B \subset A \Rightarrow A = B$

j) $A \subset B, C \subset D \Rightarrow$ $\quad$ *ja*) $A+C \subset B+D$ $\quad$ *jb*) $A.C \subset B.D$

2.2. Dokažte, že jevy $\overline{A}, A.B, A.\overline{B}$ tvoří úplnou skupinu disjunktních jevů.

2.3. Dokažte, že $\overline{A}.B + A.\overline{B} + \overline{A}.\overline{B} = \overline{A.B}$.

2.4. Dokažte, že $\overline{\overline{A}.\overline{B}} = A + B, \overline{\overline{C}+\overline{D}} = C.D$.

2.5. Dokažte ekvivalentnost a pravdivost tvrzení:

$$\overline{\sum_{k=1}^{n} A_k} = \prod_{k=1}^{n} \overline{A_k}, \ \sum_{k=1}^{n} \overline{A_k} = \overline{\prod_{k=1}^{n} A_k} .$$

2.6. Zjednodušte $A = \ B+C \ . \ B+\overline{C} \ . \ \overline{B}+C$.

2.7. Nechť $A \subset B$. Zjednodušte výrazy: *a*) $A.B$, *b*) $A + B$, *c*) $A.B.C$

2.8. Dokažte, že jev $A+B \ . \ A+\overline{B} \ . \ \overline{A}+B \ . \ \overline{A}+\overline{B}$ není možný.

2.9. A, B, C jsou náhodné jevy. Zjednodušte výrazy:

$a)\ \ A+B\ .\ B+C \quad b)\ \ A+B\ .\ A+\overline{B}\ .$

2.10. Kdy jsou možné rovnosti: $a)\ A+B=\overline{A}$, $b)\ A\cdot B=\overline{A}$, $c)\ A + B = A.B$?

2.11. Jsou jevy $A, \overline{A+B}$ disjunktní?

2.12. Dokažte, že jevy $A, B, \overline{A+B}$ tvoří úplnou skupinu vzájemně neslučitelných jevů.

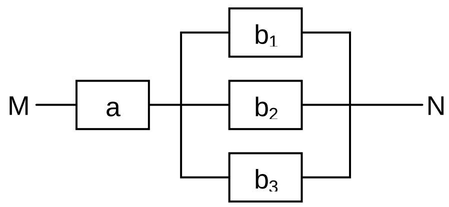
2.13. Najděte jev X z rovnice $\overline{X+A}+\overline{X+\overline{A}}=B$.

2.14. Terč je tvořen deseti kruhy ohraničenými soustřednými kružnicemi o poloměrech r_k, $\underline{k} = 1, ..., , 10$, přičemž $r_1 < r_2 < ... < r_{10}$. Určete, co značí jevy:

$a)\ \ B=\sum_{k=1}^{6} A_k\ ,\ \ b)\ \ C=\prod_{k=5}^{10} A_k\ .$

2.15. Jev A značí, že alespoň jeden ze tří výrobků, procházejících kontrolou, je vadný. Jev B značí, že všechny tři kontrolované výrobky jsou dobré. Co značí jevy $A + B$, $A\ .\ B$?

2.16. Mezi body M a N jsou zapojeny prvky a, b_1, b_2, b_3 podle schématu. Jev A značí poruchu prvku a, jev B_k poruchu prvku b_k , $k = 1, 2, 3$. Vyjádřete jevy C a $\overline{C}$ pomocí A, B_k, když C značí přerušení spojení mezi body M a N.



2.17. Přístroj se skládá ze dvou bloků 1. typu a tří bloků 2. typu.

Jevy: A_k , $k = 1, 2$ -- funguje k-tý blok 1. typu

B_j , $j = 1, 2, 3$ -- funguje j-tý blok 2. typu.

Přístroj je schopen pracovat, když funguje aspoň jeden blok 1. typu a aspoň dva bloky 2. typu. Vyjádřete jev C značící, že přístroj je v pořádku.

2.18. Při hodu hrací kostkou značí jev A "padnutí sudého čísla", jev B "padnutí čísla dělitelného 3". Určete, co znamená jev: $A + B$, $A - B$, $A\ .\ B$, $\overline{A}$, $\overline{B}$, $B - A$.

2.19. Jev A znamená, že z 10-ti automobilů byly prodány:

$a)$ alespoň 3

$b)$ alespoň 5

c) žádný

d) právě 4

e) aspoň 6 a nejvýše 8

f) žádný nebo alespoň 3

Kolik automobilů bylo prodáno, jestliže nastal jev $\overline{A}$?

2.20. Ke zkoušce jde 10 studentů. Jev A_k znamená: zkoušku udělalo alespoň k studentů. Jev B_k znamená: zkoušku udělalo nejvýše k studentů. Jev C_k znamená: zkoušku udělalo právě k studentů. Kolik studentů udělalo zkoušku, nastaly-li jevy: $A_2 \,.\, A_3$, $A_2 + A_3$, $\overline{C_3}$, $\overline{C_6}$, $B_2 \,.\, B_4$, $B_2 + B_4$, $A_2 \,.\, B_3$, $A_8 + B_2$.

2.21. Zapište pomocí symboliky uvedené v předchozím příkladě jevy:

a) zkoušku udělali 2 až 3 nebo 3 až 4 studenti

b) zkoušku udělali nejvýše 4 nebo alespoň 7 studentů

2.22. Student udělá zkoušku (jev A), jestliže napíše úspěšně písemku (jev B) a zodpoví při ústní zkoušce alespoň jednu ze tří otázek (jevy C_1, C_2, C_3). Vyjádřete jev A pomocí jevů B, C_1, C_2, C_3.

Klasická definice pravděpodobnosti

2.23. Číslice 1, 2, 3, 4, 5 jsou napsány na 5-ti lístcích. Náhodně vybereme 3 a utvoříme z nich trojciferné číslo, přičemž cifry k sobě skládáme v pořadí v jakém jsme je vybrali. Vypočtěte pravděpodobnost, že vzniklé trojciferné číslo bude sudé.

2.24. Kruhový terč má 3 pásma. Pravděpodobnost zásahu 1. pásma je 0,2, druhého 0,23 a třetího 0,15. Jaká je pravděpodobnost minutí cíle?

2.25. S jakou pravděpodobností padne na dvou kostkách součet

a) šest

b) menší než 7

2.26. Máme 230 výrobků, mezi nimiž je 20 nekvalitních. Vybereme 15 výrobků, přičemž vybrané výrobky nevracíme zpět. Jak je pravděpodobné, že mezi 15 vybranými bude 10 dobrých?

2.27. V zástupu 7 lidí jsou 3 ženy. Jaká je pravděpodobnost, že ženy stojí bezprostředně za sebou?

2.28. Do kolony bylo náhodně seřazeno 7 aut. 2 Mercedesy, 3 Hondy a 2 Oply. Jaká je pravděpodobnost, že na prvním a posledním místě bude Honda?

2.29. V osudí jsou 4 černé a 6 modrých koulí. Náhodně vybereme 4. Jaká je pravděpodobnost, že

a) 3 budou modré a jedna černá?

b) alespoň 3 vytažené koule budou modré?

c) mezi vytaženými koulemi je více černých

2.30. V telefonním seznamu náhodně vybereme jedno šestimístné číslo (může začínat nulou) a předpokládáme, že v seznamu jsou použita všechna šestimístná čísla. Jaká je pravděpodobnost, že číslo

a) neobsahuje 0

b) obsahuje jednu 3

2.31. Házíme současně třemi hracími kostkami a sčítáme bodové hodnoty. Který ze součtů 11 nebo 12 je pravděpodobnější?

Geometrická definice pravděpodobnosti

2.32. Hodiny, které nebyly ve stanovenou dobu nataženy, se po určitém čase zastaví. Jaká je pravděpodobnost, že se velká ručička zastaví mezi 6 a 9?

2.33. Tyč délky 10m je náhodně rozlomena na 2 části. Jaká je pravděpodobnost, že menší část bude delší než 4m?

2.34. Z intervalu $\langle 0,1 \rangle$ byla náhodně vybrána 2 čísla x a y. Nechť jev A značí, že $y \le x$ a jev B, že $x \le 0,5$. Určete pravděpodobnost jevů: A, B, $A.B$, $A + B$.

2.35. Na zastávku místní dopravy přijíždí autobus každých 7 minut a zdrží se 0,5 minuty. Jaká je pravděpodobnost, že přijdu a zastihnu autobus na zastávce?

2.36. Z intervalu $\langle 0,8 \rangle$ náhodně vybereme čísla x a y. Jaká je pravděpodobnost, že $y \le x^3$?

2.37. Určete pravděpodobnost toho, že součet náhodně zvolených kladných pravých zlomků není větší než jedna a současně jejich součin není větší než $\frac{2}{9}$.

2.38. Autobus přijíždí na zastávku každé 4 minuty, tramvaj (má zastávku vedle) každých 6 minut. Určete pravděpodobnost, že se cestující dočká:

a) autobusu před tramvají

b) autobusu nebo tramvaje v průběhu 2 minut

2.39. Pacient se léčí doma a od 7 do 20 hod. je možné jej kontrolovat. Vycházky má od 13 do 15 hod. Jaká je pravděpodobnost, že mezi 7. a 20. hodinou bude doma k zastižení?

Podmíněná pravděpodobnost

2.40. Házíme dvěma kostkami. Vypočtěte, jaká je pravděpodobnost toho, že:

a) padne-li na 1.kostce dvojka, padne součet větší než 6.

b) padne-li na 1. kostce sudé číslo, padne součet větší než 8.

2.41. Z celkové produkce závodu jsou 4 % zmetků a z dobrých je 75 % standardních. Určete pravděpodobnost, že náhodně vybraný výrobek je standardní.

2.42. Z výrobků určitého druhu dosahuje 95 % předepsanou kvalitu. V určitém závodě, který vyrábí 80 % celkové produkce však předepsanou kvalitu má 98 % výrobků. Mějme náhodně vybraný výrobek předepsané kvality. Jaká je pravděpodobnost, že byl vyroben ve výše uvedeném závodě?

2.43. V zásilce je 90 % standardních výrobků, mezi nimiž je 60 % výrobků mimořádné kvality. Vypočítejte jaká je pravděpodobnost, že náhodně vybraný výrobek z celé zásilky je mimořádně kvalitní.

2.44. Tři závody vyrábí žárovky. První 45 % celkové produkce, druhý 40 % a třetí 15 %. Z produkce prvního závodu je standardních 70 %, druhého 80 % a třetího 81 %. Určete pravděpodobnost, že si zákazník koupí standardní žárovku.

2.45. Menza VŠB zakoupila 12 chladniček z 1. závodu, 20 z 2. závodu a 18 z 3. závodu. Pravděpodobnost, že chladnička je výborné jakosti, pochází-li z 1. závodu je 0,9, z 2. závodu 0,6 a z 3. závodu 0,9. Jaká je pravděpodobnost, že náhodně vybraná chladnička bude výborné jakosti?

2.46. Součástky, ze kterých se montují stroje, dodávají tři závody. Je známo, že první má

0,3 % zmetků, druhý 0,2 % zmetků a třetí 0,4 %. Přitom první závod dodal 1000, druhý 2000 a třetí 2500 součástek. Jaká je pravděpodobnost, že náhodně vybraná součástka bude zmetek?

2.47. Máme 4 krabice. V první jsou 3 bílé a 2 černé koule, ve druhé jsou 2 bílé a 2 černé koule, ve třetí je 1 bílá a 4 černé koule, ve čtvrté 5 bílých a 1 černá koule. Náhodně vybereme jednu krabici a vytáhneme 1 kuličku. Jaká je pravděpodobnost, že kulička je bílá?

2.48. Ve společnosti je 45 % mužů a 55 % žen. Vysokých nad 190 cm je 5 % mužů a 1 % žen. Náhodně vybraná osoba je vyšší než 190 cm. Jaká je pravděpodobnost, že je to žena?

2.49. V dílně pracuje 10 dělníků, kteří vyrobí za směnu stejný počet výrobků. Pět z nich vyrobí 96 % standardních, tři z nich 90 % standardních a dva 85 % standardních. Všechny výrobky jdou do skladu. Náhodně jsme vybrali jeden výrobek a zjistili, že je standardní. Jaká je pravděpodobnost, že ho vyrobil někdo z prvních pěti dělníků?

Opakované pokusy

2.50. V populaci se vyskytují 4 % homosexuálně zaměřených jedinců. Jaká je pravděpodobnost, že ve 20-ti členné studijní skupině bude alespoň jeden takto zaměřený jedinec?

2.51. Dva sportovní střelci nezávisle na sobě střílejí do jednoho terče. Každý po jednom výstřelu. Pravděpodobnost zásahu prvního střelce je 0,8, druhého 0,4. Při střelbě byl v terči jeden zásah. Jaká je pravděpodobnost, že terč zasáhl první střelec?

2.52. Sportovní střelec zasáhne cíl při každém výstřelu s pravděpodobností p = 0,8. Vypočtěte pravděpodobnost, že při 5 výstřelech budou v cíli

a) právě 2 zásahy,

b) nejvýše jeden zásah,

c) alespoň 2 zásahy.

2.53. Určete pravděpodobnost, že při pěti hodech kostkou padne:

a) šestka právě dvakrát,

b) šestka při druhém a čtvrtém hodu.

2.54. Písemná zkouška z matematiky obsahuje 5 příkladů. Pravděpodobnost spočítání jednoho příkladu je 0,8. Určete, jaká je pravděpodobnost, že student uspěje, stačí-li, aby spočítal aspoň 3 příklady.

2.55. V rodině je *n* dětí. Pravděpodobnost narození chlapce je 0,515. Určete počet dětí tak, aby mezi nimi byl aspoň jeden chlapec s pravděpodobností alespoň 0,99.

2.56. Pravděpodobnost výhry hráče je 0,6. Určete, jaký je nejpravděpodobnější počet výher hráče v deseti odehraných partiích.

2.57. Sérii 100ks výrobků je třeba zkontrolovat náhodným výběrem. Celá je považována za špatnou, je-li aspoň jeden z pěti vybraných výrobků vadný. Vypočtěte pravděpodobnost, že série je špatná, víme-li, že obsahuje 5 % vadných výrobků.

Úlohy k samostatnému řešení - netříděno

2.58. Máme dřevěnou krychli, jejíž stěny jsou červeně obarveny. Rozřežme ji na 125 stejných krychliček, které vzájemně promícháme. Potom náhodně vybereme jednu krychličku. Jaká bude pravděpodobnost, že vybraná krychlička bude mít dvě stěny červeně natřené?

2.59. V jedné studijní skupině prvého ročníku FAST v Brně je 24 posluchačů, z nichž 5 má trvalé bydliště v Brně, 6 v Ostravě a zbývající jsou odjinud. Na výrobní praxi do Ostravy bylo ze skupiny namátkou vybráno 12 posluchačů. Jaká je pravděpodobnost, že mezi vybranými budou

a) všichni posluchači z Ostravy,

b) 3 posluchači z Ostravy,

c) žádný posluchač z Ostravy.

2.60. Ke kontrole je připravena skupina 200 výrobků, z nichž jsou 4 % vadných. Ostatní mají požadovanou kvalitu. Namátkou z nich vybereme 20 kusů. Při kontrole zjišťujeme, že prvních 5 z 20 vybraných je kvalitních. Jaká je pravděpodobnost, že šestý výrobek je též kvalitní?

2.61. Máme karetní hru o 32 kartách. Vytáhneme jednu kartu, vrátíme ji a karty promícháme. Potom znovu vytáhneme jednu kartu. Určete pravděpodobnost toho, že obě karty budou stejné barvy.

2.62. Na deseti stejných kartičkách jsou čísla od nuly do devíti. Určete pravděpodobnost toho, že dvojmístné číslo (může začínat nulou) náhodně vytvořené z daných kartiček je dělitelné

a) 6,

b) 21.

2.63. Karetní hru o 52 kartách dělíme libovolně na dvě stejné části. Jaká je pravděpodobnost, že v každé části budou dvě esa?

2.64. Z karetní hry o 32 kartách náhodně vybereme 3 karty. Jaká je pravděpodobnost, že mezi nimi bude aspoň jeden král?

2.65. V osudí je 5 koulí bílých a 5 černých. Vybíráme bez vracení 6 koulí. Jaká je pravděpodobnost, že

a) dvě koule z vybraných budou bílé,

b) alespoň dvě koule z vybraných budou bílé?

2.66. V osudí je 8 koulí bílých a 6 červených. Vybereme náhodně 4 koule. Jaká je pravděpodobnost, že vybrané koule nejsou všechny stejné barvy.

2.67. V laboratoři se má zjistit mez průtažnosti vzorku oceli. Pravděpodobnost toho, že mez průtažnosti bude v rozmezí 27-29 kp/mm^2, je 0,14; pro rozmezí 29-31 kp/mm^2 je pravděpodobnost 0,21; pro rozmezí 31-33 kp/mm^2 je 0,16. Určete, jaká je pravděpodobnost toho, že mez průtažnosti zkoumaného vzorku je v rozmezí 27-33 kp/mm^2.

2.68. Výrobek prochází v průběhu zpracování postupně čtyřmi operacemi. Pravděpodobnost vyrobení zmetku je u jednotlivých operací postupně rovna 0,02; 0,03; 0,005; 0,015. Určete přibližně pravděpodobnost toho, že výsledkem výrobního procesu v daném případě bude zmetek.

2.69. Vytočíme náhodně pěticiferné telefonní číslo. Jaká je pravděpodobnost, že vytočíme buď číslo 31540 nebo číslo 71432, víme-li, že telefonní číslo bude mít jako prvou číslici některou z cifer 3, 5, 7, 9?

2.70. Pět žárovek ze sta se namátkou kontroluje. Při výběru žárovky nevracíme. Vyskytne-li se mezi pěti kontrolovanými zmetek, je celá stovka vyřazena jako zmetkovitá. Jaká je pravděpodobnost, že daných sto žárovek bude vyřazeno, víme-li, že je mezi nimi 6 zmetků?

2.71. Z n výrobků, v nichž je r zmetků, náhodně bereme bez vracení r výrobků. Jaká je pravděpodobnost toho, že vybereme všechny zmetky?

2.72. V osudí je n lístků s čísly od 1 do n. Lístky vytahujeme po jednom bez vracení. Jaká je pravděpodobnost toho, že při prvých k tazích budou čísla na lístcích stejná jako počet provedených tahů?

2.73. Házíme čtyřikrát hrací kostkou. Jaká bude pravděpodobnost, že při každém hodu dostaneme jiný počet oček?

2.74. Z osudí, v němž je n koulí, n-krát vytáhneme kouli a vždy ji vrátíme zpět. Jaká je pravděpodobnost, že postupně vyjmeme všechny koule?

2.75. Studijní skupina, v níž je 6 studentek a 18 studentů, se pro laboratorní cvičení náhodně rozděluje na 6 skupin po čtyřech. Jaká je pravděpodobnost, že v každé skupině bude studentka?

2.76. Házíme dvakrát kostkou. Jaká je pravděpodobnost, že podruhé padne více oček než poprvé?

2.77. Dva závodníci zdolají určitou vzdálenost ve stanoveném čase s pravděpodobností 0,8 a 0,9. Určete pravděpodobnost, že ve stanoveném čase dosáhne cíle alespoň jeden závodník.

2.78. Z osudí, v němž je 10 koulí bílých a 2 červené, táhneme n-krát po jedné kouli a po každém tahu ji vrátíme zpět. Určete nejmenší hodnotu n tak, aby pravděpodobnost jevu, že alespoň jednou vytáhneme červenou kouli, byla větší než 1/2.

2.79. Z osudí, v němž je 12 koulí bílých a 2 červené, táhneme m-krát bez vracení. Určete nejmenší hodnotu m tak, aby pravděpodobnost jevu, že alespoň jednou vytáhneme červenou kouli, byla větší než 1/2.

2.80. Kolikrát musíme hodit třemi kostkami, aby pravděpodobnost jevu, že alespoň jednou padne 18 ok, byla větší než 1/2?

2.81. Dva hráči házejí mincí. Vyhrává ten, komu dřív padne líc. Určete pravděpodobnost výhry každého hráče.

2.82. Dva střelci postupně střílejí na cíl do prvého zásahu. Pravděpodobnost zásahu pro prvého střelce je 0,2, pro druhého 0,3. Určete pravděpodobnost toho, že první střelec bude mít více výstřelů než druhý.

2.83. Tři rovnocenní hráči A,B,C hrají společenskou hru. Určete, zda je pravděpodobnější, že hráč A vyhraje 3 ze 4 nebo 5 z 8 partií.

2.84. V osudí je 10 koulí - 3 bílé a 7 černých. Pětkrát táhneme po jedné kouli, po každém tahu ji vrátíme zpět. Určete pravděpodobnost, že budou taženy buď všechny koule bílé, nebo všechny černé.

2.85. Pravděpodobnost toho, že jev A nastane při jednom pokusu, je p. Určete pravděpodobnost nastoupení téhož jevu alespoň jednou při pěti pokusech.

2.86. V osudí je 5 lístků s čísly od 1 do 20. Provedeme a) 3 tahy, b) 5 tahů. Po každém tahu lístek vrátíme zpět a lístky znovu zamícháme. Určete pravděpodobnost toho, že v každém z obou uvedených případů alespoň 2-krát vytáhneme lístek s číslem dělitelným čtyřmi.

2.87. Házíme pětkrát hrací kostkou. Určete pravděpodobnost toho, že alespoň ve dvou hodech, ale zároveň ne víc jak čtyřikrát, padne počet ok dělitelný třemi.

2.88. Z karetní hry o 32 kartách 20-krát táhneme po jedné kartě, po každém tahu kartu vrátíme zpět. Určete nejpravděpodobnější počet tahů x_0, v nichž se nám podaří vytáhnout eso, a pro vypočtené x_0 určete příslušnou pravděpodobnost.

2.89. Pravděpodobnost toho, že množství odebraného elektrického proudu v určitém závodě je normální (nepřesáhne plánovanou spotřebu za 24 hod.), je rovna 3/4. Stanovte pravděpodobnost, že v nejbližších šesti dnech bude alespoň po dobu tří dnů odběr proudu normální.

2.90. Pravděpodobnost toho, že v některém okamžiku během jednoho roku bude na určitou konstrukci působit současně maximální zatížení pohyblivé a maximální zatížení větrem, činí 3.10^{-8}. Tato pravděpodobnost se během let nemění. Životnost konstrukce je 100 let. Jaká je pravděpodobnost, že za dobu trvání konstrukce se obě zatížení ve svých maximálních hodnotách střetnou alespoň jednou?

2.91. Pravděpodobnost toho, že mužstvo A vyhraje aspoň jedno ze čtyř utkání, je rovna 0,59. Určete pravděpodobnost vítězství mužstva A v jednom utkání, předpokládáme-li že všichni čtyři soupeři jmenovaného mužstva mají stejnou úroveň.

2.92. Na dvojkolejním železničním mostě se potkají v průběhu 24 hodin dva protijedoucí vlaky s pravděpodobností 0,2. Určete pravděpodobnost toho, že v průběhu týdne se dva vlaky na mostě potkají

a) maximálně třikrát,

b) nejméně třikrát,

c) právě třikrát.

d) Určete, kolikrát se vlaky potkají s největší pravděpodobností.

2.93. Pravděpodobnost toho, že televizní obrazovka vydrží bez poruchy 3000 hodin provozu, je 0,4.

a) Jaká je pravděpodobnost toho, že alespoň jedna z pěti stejných obrazovek vydrží bez poruchy 3000 hodin?

b) Jaký nejpravděpodobnější počet z pěti obrazovek vydrží stanovený počet hodin bez poruchy?

2.94. Na nosník délky L umístíme libovolně dvě břemena. S jakou pravděpodobností je umístíme tak, že jejich vzdálenost

a) nebude větší než L/4,

b) nebude větší než L/2?

2.95. Dva lidé se dohodli, že se setkají na stanoveném místě mezi 18:00 h. a 18:45 h. Ten, kdo přijde první, počká na druhého 15 minut. Určete pravděpodobnost toho, že se setkají, je-li příchod obou kdykoliv ve stanoveném čase stejně možný.

2.96. Stanovte pravděpodobnost toho, že výraz

$$z = \sqrt{\frac{x^2 + y^2}{x.y - 1}}$$

je v libovolném bodě *(x, y)* definován, může-li *x* a *y* nabýt se stejnou pravděpodobností libovolné hodnoty z oboru $|x| \le 2, |y| \le 2$.

2.97. Určete pravděpodobnost, s jakou bude v libovolném bodě oblasti $x \in \langle -1; 2\rangle \wedge |y| < 2$ definována funkce $z = \ln \quad -x - y$.

2.98. Určete pravděpodobnost toho, že libovolně zvolený bod uvnitř krychle o hraně 10, jejíž střed leží v počátku a hrany jsou rovnoběžné s osami souřadnými, je současně bodem definičního oboru funkce

$$u = \sqrt{9 - x^2 - y^2 - z^2} + \frac{1}{\sqrt{x^2 + y^2 + z^2 - 4}}$$

.

2.99. Mějme terč tvořený dvěma soustřednými kružnicemi o poloměrech $2r$ a $3r$. Předpokládáme stejnou pravděpodobnost zásahu do libovolného bodu terče. Určete pravděpodobnost toho, že ze tří zásahů terče bude jeden zásah do vnitřního kruhu.

2.100. Na úsečce délky L jsou náhodně zvoleny dva body, čímž je tato úsečka rozdělena na tří části. Určit pravděpodobnost toho, že z těchto tří úseček je možno sestrojit trojúhelník.

2.101. Na kružnici o poloměru R jsou náhodně zvoleny body A, B, C. Jaká je pravděpodobnost, že trojúhelník ABC je ostroúhlý?

2.102. Na stavbu byly dovezeny cihly ze tří cihelen a složeny na společné skládce. Jejich množství jsou v poměru 1:2:2. Cihly vyrobené jednotlivými cihelnami vyhoví předepsaným normám jakosti s pravděpodobností rovnou postupně 0,80, 0,65, 0,72. Ze skládky cihel náhodně vybereme jeden kus, abychom laboratorně zjistili, zda splňuje předepsané požadavky. Jaká je pravděpodobnost toho, že cihla bude mít předepsanou kvalitu?

2.103. V osudí je 24 koulí - 4 černé, 12 červených a 8 bílých. Určete pravděpodobnost, že v druhém tahu vytáhneme bílou kouli, nevíme-li, jakou kouli jsme vytáhli v 1. tahu. Koule do osudí nevracíme.

2.104. Máme u schránek, v nichž je v každé m bílých a n šedých stejně velkých obálek. Z prvé schránky náhodně vybereme obálku a vložíme ji do druhé. Z druhé opět vytáhneme jednu obálku a vložíme ji do třetí, atd. Určete pravděpodobnost toho, že po takovém přemístění vytáhneme z poslední schránky bílou obálku.

2.105. Do urny, v níž je n koulí, je vhozena bílá koule. S jakou pravděpodobností je pak možno z urny vytáhnout bílou kouli, když všechny předpoklady o původním stavu v urně jsou stejně pravděpodobné?

2.106. Máme čtyři osudí. V prvém jsou 3 koule bílé a 2 černé, v druhém a třetím po 2 bílých a 5 černých, ve čtvrtém je 1 bílá a 3 černé koule. Můžeme předpokládat, že vytažení koule z libovolného osudí je stejně pravděpodobné. Určete pravděpodobnost, že *a*) vytažená bílá koule je z prvé urny,

b) vytažená černá koule je ze čtvrté urny.

2.107. K síti je připojeno 14 nových a 6 starších počítačů. Pravděpodobnost bezchybného provozu u nových počítačů je 0.9, u starších 0.8. Jaká je pravděpodobnost, že

a) student bude pracovat bez poruchy

b) tento student pracuje u nového počítače?

2.108. Házíme třikrát hrací kostkou. Najděte pravděpodobnost následujících jevů:

A - na všech kostkách padnou tři oka

B - na všech kostkách padne týž počet ok

C - na kostkách padnou různé počty ok

2.109. Do výtahu v sedmipodlažním domě nastoupili v 1. podlaží tři lidé. Každý z nich se stejnou pravděpodobností může vystoupit v libovolném podlaží počínaje druhým. Najděte pravděpodobnost následujících jevů:

A - všichni cestující vystoupí ve čtvrtém podlaží

B - všichni cestující vystoupí současně

C - cestující vystoupí v různých podlažích

Výsledky úloh k samostatnému řešení

2.6. $A = B\,C$

2.7. *a)* A

b) B

c) $A\,C$

2.9. *a)* $B + A\,C$ b) A

2.10. *a)* $A = \emptyset, B = I$

b) $A = I, B = \emptyset$

c) $A = B$

2.11. ano

2.13. $X = \overline{B}$

2.14. *a)* $B = A_6$

b) $C = A_5$

2.15. $A + B = I$, $A.B = \varnothing$

2.16. $C = A + B_1\, B_2\, B_3$

$\overline{C} = \overline{A}.\ \overline{B_1} + \overline{B_2} + \overline{B_3}$

2.17. $C = (A_1 + \mathrm{A}_2)\,(B_1\, B_2 + B_2\, B_3 + B_1\, B_3)$

2.18. A+B... padne 2 nebo 3 nebo 4 nebo 6

A-B... padne 2 nebo 4

$A.B$... padne 6

$\overline{A}$... padne 1 nebo 3 nebo 5

$\overline{B}$... padne 1 nebo 2 nebo 4 nebo 5

B-A... padne 3

2.19. *a*) nejvýše 2

b) nejvýše 4

c) aspoň 1 *d*) nejvýše 3 nebo aspoň 5

e) nejvýše 5 nebo aspoň 9

f) jeden nebo dva

2.20. $A_2.A_3 = A_3$

$A_2+A_3 = A_2$

$\overline{C_3} = B_2+A_4$

(nejvýše 2 nebo aspoň 4)

$\overline{C_6} = B_5+A_7$

(nejvýše 5 nebo aspoň 7)

$B_2.B_4 = B_2$

$B_2+B_4 = B_4$

$A_2.B_3 = C_2+C_3$(2 nebo 3)

$A_8+B_2 = C_0+C_1+C_2+C_8+C_9+C_{10}$

(nejvýše 2 nebo alespoň 8)

2.21. *a*) $A_2.B_3+A_3.B_4$

b) B_4+A_7

2.22. $A = B.(C_1+C_2+C_3)$

2.23. 0,4

2.24. 0,42

2.25. 0,1388; 0,4166

2.26. 0,004

2.27. 0,142

2.28. 0,142

2.29. 0,38; 0,452; 0,119

2.30. 0,531; 0,354

2.31. 11

2.32. 0,25

2.33. 0,2

2.34. 0,5; 0,5; 0,125; 0,875

2.35. 0,07

2.36. 0,812

2.37. 0,487

2.38. 0,66; 0,66

2.39. 0,846

2.40. 0,33; 0,33

2.41. 0,72

2.42. 0,825

2.43. 0,54

2.44. 0,7565

2.45. 0,78

2.46. 0,003

2.47. 0,53

2.48. 0,196

2.49. 0.52

2.50. 0,558

2.51. 0,857

2.52. 0,0512; 0,0067; 0,9932

2.53. 0,16; 0,016

2.54. 0,942

2.55. 7

2.56. 6

2.57. 0,2305

2.58. 0,288

2.59. *a*) $C_6(6)*C_6(18) / C_{12}(24)$= 0,00686498

b)$C_3(6)*C_9(18) / C_{12}(24)$= 0,359594

c) $C_0(6)*C_{12}(18) / C_{12}(24)$ = 0,00686498

2.60. 187 / 195 = 0,958974

2.61. 32 / 32 * 8 / 32 = 0,25

2.62. *a*) 15 / 90

b) 4 / 90

2.63. $C_2(4)*C_{24}(48) / C_{26}(52)$ = 0,390156

2.64. 1 - $C_3(28) / C_3(32)$ = 0,339516

2.65. *a*) $C_2(5) * C_4(5) / C_6(10)$

b) $(C_2(5)*C_4(5)+C_3(5)*C_3(5)+$

$+C_4(5)*C_2+C_5(5)*C_5(5))/ C_6(10)$ =

= 1 - $C_5(1)*C_5(5)/C_6(10)$ = 0,976190

2.66. $1 - (C_4(8) / C_4(14) + C_4(6) / C_4(14)) = 0{,}915084$

2.67. 0,51

2.68. 1 - 0,98 * 0,97 * 0,995 * 0,985 = 0,0683407

2.69. 0,00005

2.70. 1 - 94/100 * 93/99 * 92/98 * 91/97 * 90/96 =
$= 1 - C_5(94) / C_5(100) = 0{,}270914$

2.71. $r/n*(r-1)/(n-1)*...*1/(n-(r-1)) = 1 / C_r(n)$

2.72. $1/n*1/(n-1)*...*1/(n-(r-1) = 1/V_k(n) = 1 / (C_k(n)*k!)$

2.73. 6/6 * 5/6 * 4/6 * 3/6 = 5 / 18 = 0,277777

2.74. $n/n * (n-1)/n *...*1/n = n! / n^n$

2.75. $C_1(6)C_3(18)/C_4(24)*C_1(5)*C_3(15)/C_4(20)*C_1(4)*C_3(12)/C_4(16)*$
$*C_1(3)*C_3(9)/C_4(12)*C_1(2)*C_3(6)/C_4(8)*C_1(1)*C_3(3)/C_4(4) = 0{,}0304318$

2.76. 1/6*5/6+1/6*4/6+1/6*3/6+1/6*2/6+1/6*1/6 = 0,41666666

2.77. 1 - (1-0,8)*(1-0,9) = 0,98

2.78. $1 - (5/6)^n > 1/2$; $n_{\min} = 4$

2.79. $1 - C_m(12) / C_m(14) > 1/2$; $m = 4$

2.80. $1 - (215 / 216)^n > 1/2$; $n \geq 150$

2.81. $p(A)=1/2+1/2*1/2*1/2+...+1/(2^{(n-1)}-1)*2) = 2/3$
$p(B)=1/2*1/2+1/2*1/2*1/2*1/2+...+1/(2^2*2^n) = 1/3$

2.82. $p_1+q_1*q_2*p_1+...+(q_1*q_2)^{(n-1)}*p_1=p_1(1-q_1*q_2) = 5/11$

2.83. $p_{3/4}=C_3(4)*(1/3)*(2/3)=8/11=0{,}0987654$
$p_{5/8}=C_5(8)*(1/5)^5*(2/3)^3= 448/6581=0{,}0682822$

2.84. $C_5(5)*(3/10)^5*(7/10)^0+C_5(5)*(7/10)^5*(3/10)^0 = 0{,}17050$

2.85. $1 - (1-p)^5$

2.86. *a*) $C_2(3)*(5/20)^2*/15/20)+C_3(3)*(1/4)^3*(15/20)^0= 0{,}15625$
b) $1-C_0(5)*(1/4)^0*(3/4)5-C_1(5)*(1/4)^1*(3/4)^4= 47/128 = 0{,}3671$

2.87. $C_2(5)*(2/6)^2*(4/6)^3+C_3(5)*(2/6)^3*(4/6)^2+C_4(2/6)^4*(4/6)^1 = 130/243 = 0{,}5349$

2.88. $C_{x-1}(n)p^{x-1}q^{n-x+1} \leq C_x(n)p^x q^{n-x} \geq C_{x+1}(\mathrm{n})p^{x+1}q^{n-x-1}$

$x_0 = 2$; $P_2(20) = C_2(20)*(1/8)2*(7/8)^{16} = 0{,}26838$

2.89. $1-(C_0(6)*(3/4)^0*(1/6)^6 + C_1(6)*(3/4)^1*(1/4)^5 + C_2(6)*(3/4)^2*(1/4)^4) = 0{,}9624$

2.90. $p(A) = (1-3*10^{-8})^{100} \approx 1 - 3*10^{-8}*100$

$p(A) = 1 - p(A) \approx 3*10^{-6}$

2.91. $0{,}59 = 1 - (1 - p)^4 \rightarrow p \approx 0{,}2$

2.92. *a*) $p(x \leq 3) = \sum C_i(7)*0{,}2^i*0{,}8^{7-i}$, $i = 0 \dots 3$

b) $p(x \geq 3) = 1 - \sum C_i(7)*0{,}2^i*0{,}8^{7-i}$, $i = 0 \dots 2$

c) $p(x=3) = C_3(7)*0{,}2^3*0{,}8^4 \approx 0{,}11469$

d) $(n+1)*p-1 \leq x \leq (n+1)*p \rightarrow x = 1$

2.93. *a*) $1 - C_0(5)*(1 - 0{,}4)^5 \approx 0{,}92224$

b) $x = 2$

2.94. x, y in $<0, L>$

a)$| x - y | \leq L/4 \rightarrow p = 7/16$

b) $| x - y | \leq L/2 \rightarrow p = 3/4$

2.95. x, y in $<0, 45>$

$| x - y | \leq 15 \rightarrow p = 5/9$

2.96. $x \,.\, y - 1 > \rightarrow y > 1/x$, $x > 0$

$y < 1/x$, $x < 0$

$p = 2 * \mathrm{int}(2 - 1/x, x, 0, 2) \approx 0.2017$

2.97. 3/8

2.98. $76\,\pi / 3000 \approx 0{,}07958$

2.99. $C_1(3) * 4/9 * (5/9)^2 \approx 0{,}411522$

2.100. 1/4

2.101. 1/4

2.102. 0,708

2.103. $8/24 * 7/23 + 16/24 * 8/23 = 1/3$

2.104. $m / (m + n)$

2.105. $1/(n+1) * (1/(n+1) + 2/(n+1) + \dots + (n+1)/(n+1)) = (n+2)/(2(n+1))$

2.106. *a*) A ... vytažení bílé $p(A) = 1/4 * (3/5 + 2/7 + 2/7 + 1/4) = 199/560$

$p(U_1/A) = (1/4*3/5)/(199/560) = 0{,}42211$

b) $(1/4*3/4)/(361/560) = 0{,}2908$

2.107. *a*) 0,870

b) 0,724

2.108. $p(A) = 1/6^3$

$p(B) = 6 / 6^3$

$p(C) = C_3(6) / 6^3$

2.109. viz výsledky příkladu 2.108.

3. NÁHODNÁ VELIČINA

Průvodce studiem

V předchozích kapitolách jste se seznámili s kombinatorikou a pravděpodobností jevů. Tyto znalosti použijeme v této kapitole, zavedeme pojem náhodná veličina, funkce, které náhodnou veličinu popisují, a číselné charakteristiky náhodné veličiny.

Předpokládané znalosti

Pojmy z pravděpodobnosti, derivace, integrál.

Cíle

Cílem této kapitoly je objasnit pojmy náhodná veličina, pravděpodobnostní funkce, hustota pravděpodobnosti, distribuční funkce, střední hodnota, rozptyl, koeficient šikmosti, koeficient špičatosti, *p*-kvantil, medián, modus.

Výklad

3.1. Náhodná veličina

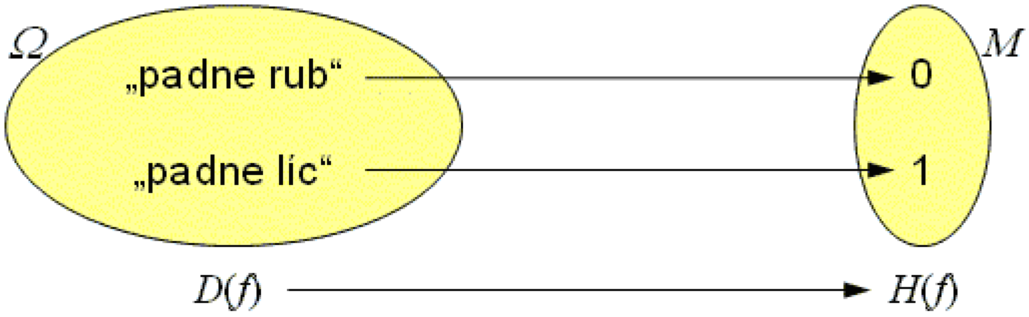
Výsledky některých pokusů (elementární jevy) jsou přímo vyjádřeny číselně (padne 1), u jiných tomu tak není (padne líc). Také u těchto pokusů je účelné přiřadit elementárním jevům čísla.

Čísla přiřazená elementárním jevům tvoří **obor hodnot *M*** proměnné, kterou nazýváme **náhodná veličina** (označujeme X, Y, Z,...)

Definice 3.1.1.
Náhodná veličina *X* je reálná funkce definovaná na množině všech elementárních jevů, která každému jevu přiřadí reálné číslo.

Např.:

Hod mincí



Podle oboru hodnot M rozdělujeme náhodné veličiny na:

- diskrétní . . . obor hodnot M je konečná nebo nekonečná posloupnost
- spojité . . . obor hodnot M je otevřený nebo uzavřený interval

3.2. Diskrétní náhodná veličina

3.2.1. Pravděpodobnostní funkce

Nechť X je diskrétní náhodná veličina s oborem možných hodnot $\{x_1, x_2, ..., e_n\}$, která tyto hodnoty nabývá s pravděpodobností $\{p_1, p_2, ..., p_n\}$.

Údaje sestavíme do tabulky:

$\boldsymbol{x_i}$	x_1	x_2	...	x_n
$\boldsymbol{p_i}$	p_1	p_2	...	p_n

Každé hodnotě x_i je přiřazena právě jedna hodnota p_i a pravděpodobnostní tabulku lze tedy chápat jako tabulkové určení funkce, kterou nazýváme **pravděpodobnostní funkcí**.

Definice 3.2.1.

Pravděpodobnostní funkcí náhodné veličiny X nazýváme funkci $p(x) = P(X = x)$

Poznámka

Funkční hodnota v x_i představuje pravděpodobnost, že náhodná veličina X nabude hodnotu x_i.

Vlastnosti pravděpodobnostní funkce:

$a)$ $p(x_i) \geq 0$

$b)$ $\sum_{i=1}^{n} p\left(x_i\right) = 1$

Poznámka

První vlastnost plyne přímo z definice pravděpodobnostní funkce. Druhé tvrzení plyne z toho, že náhodné veličině X je přiřazeno číslo x_i právě tehdy, když nastane jev s hodnotou x_i (stručněji jev X_i). Přitom jevy $X_1, X_2, \ldots, X_n$ tvoří úplnou skupinu vzájemně disjunktních jevů, protože v jednom pokusu nabývá náhodná veličina X právě jedné hodnoty z oboru M. Sečteme-li všechny možné výsledky pokusu, dostáváme jev jistý I s pravděpodobností P(I) = 1.

3.2.2. Distribuční funkce diskrétní náhodné veličiny

Často nás nezajímá jen pravděpodobnost toho, že X nabude určitou hodnotu x_i, ale potřebujeme určit pravděpodobnost, se kterou X nabude hodnoty menší než jistá mez:

Definice 3.2.2.

Reálná funkce, která přiřazuje každé hodnotě x_i náhodné veličiny X pravděpodobnost, že X nabude hodnoty menší než toto x_i, se nazývá **distribuční funkce** $F(x)$. Je definována vztahem:

$$F(x) = P(X < x) = \sum_{x_i < x} P\left(X = x_i\right)$$

Poznámka

Vlastnosti distribuční funkce budou souhrnně popsány u spojité náhodné veličiny.

Řešené úlohy

Příklad 3.2.1. Hod kostkou.

Řešení: Náhodná veličina X je definována na množině elementárních jevů: padne 1, padne 2, ..., padne 6. Obor hodnot M jsou reálná čísla $\{1,2,...,6\}$ přiřazená elementárním jevům $E_1, E_2, ..., E_6$ s pravděpodobností $\{p_1, p_2, ..., p_6\}$, kde $p_i = \frac{1}{6}$.

Pravděpodobnostní funkce $p(x) = P(X = x) = \frac{1}{6}$

Příklad 3.2.2. V osudí je 5 bílých a 7 červených míčků. Náhodná veličina X představuje počet bílých míčků mezi pěti vybranými. Vytvořte pravděpodobnostní a distribuční funkci této náhodné veličiny.

Řešení: Náhodná veličina X nabývá hodnot $\{0,1,2,3,4,5\}$.

Z teorie pravděpodobnosti víme, že se jedná o opakované závislé pokusy. Můžeme tedy sestavit pravděpodobnostní funkci:

$$p(x_i) = \frac{\binom{5}{x_i} \cdot \binom{7}{5-x_i}}{\binom{12}{5}}$$

Dosazením do pravděpodobnostní funkce vytvoříme pravděpodobnostní tabulku:

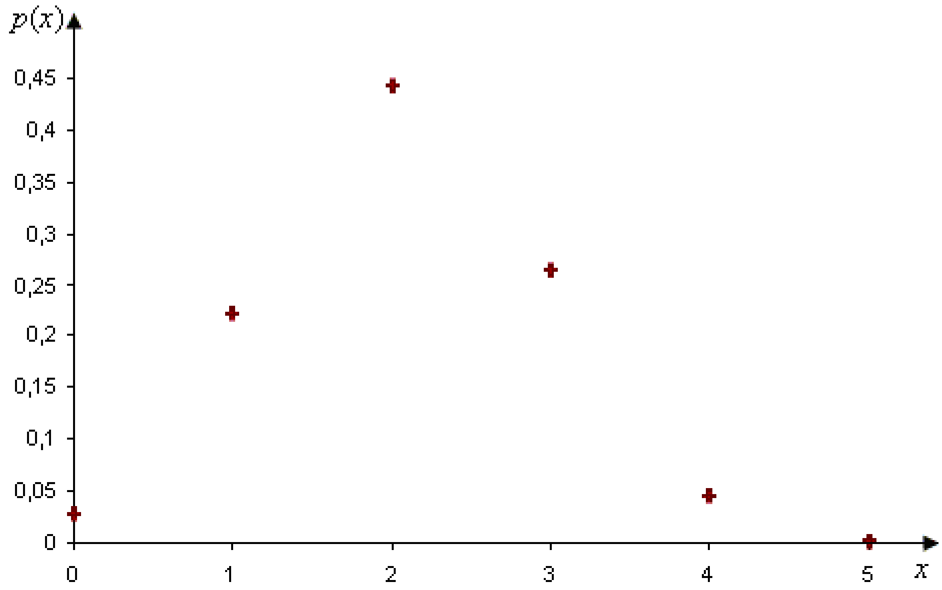
x_i	0	1	2	3	4	5
p_i	$\frac{21}{792}$	$\frac{175}{792}$	$\frac{350}{792}$	$\frac{210}{792}$	$\frac{35}{792}$	$\frac{1}{792}$

Např.

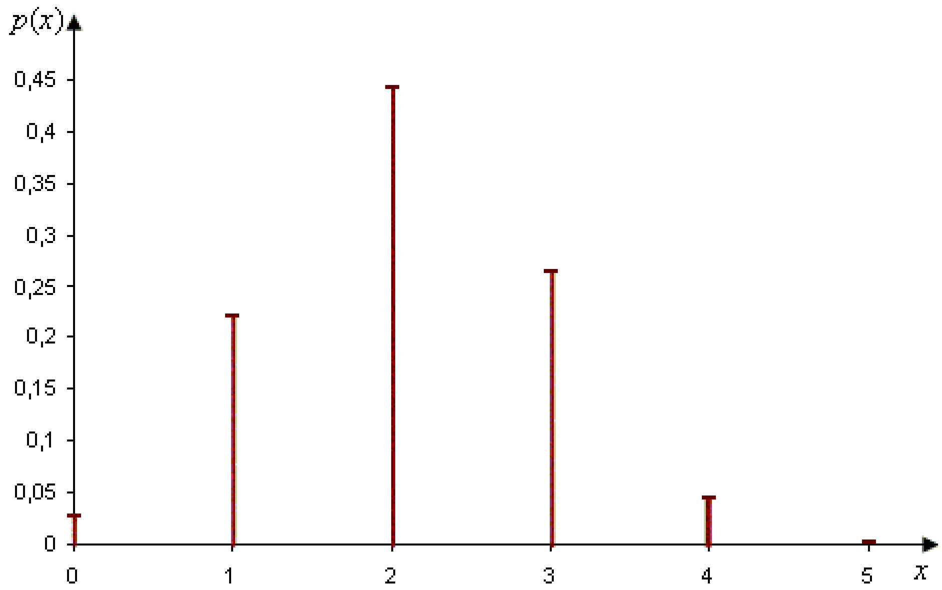
$$p_1 = p(x_1) = p(0) = \frac{\binom{5}{0} \cdot \binom{7}{5}}{\binom{12}{5}} = \frac{1.21}{792} = \frac{21}{792}$$

Možnosti grafického znázornění:

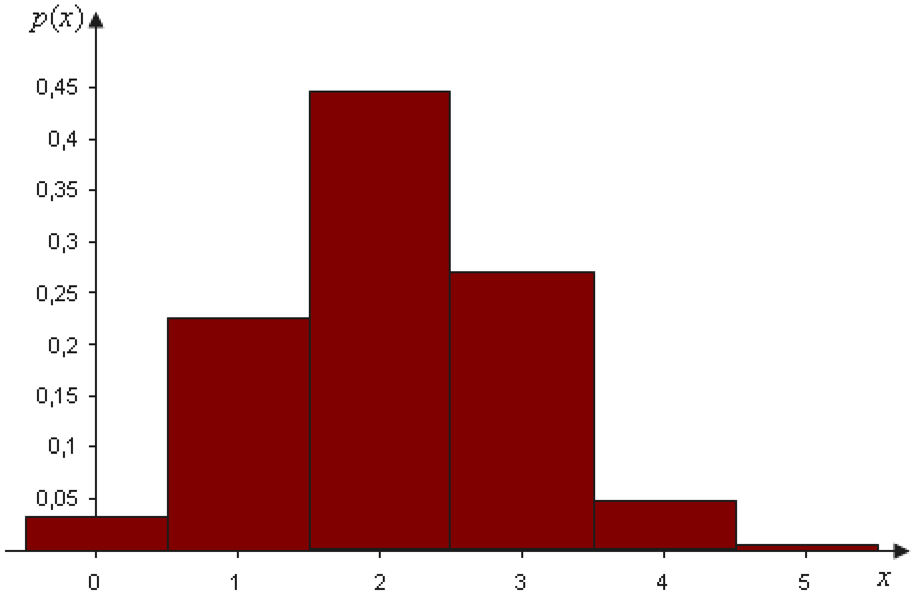
Bodový graf:



Úsečkový diagram:



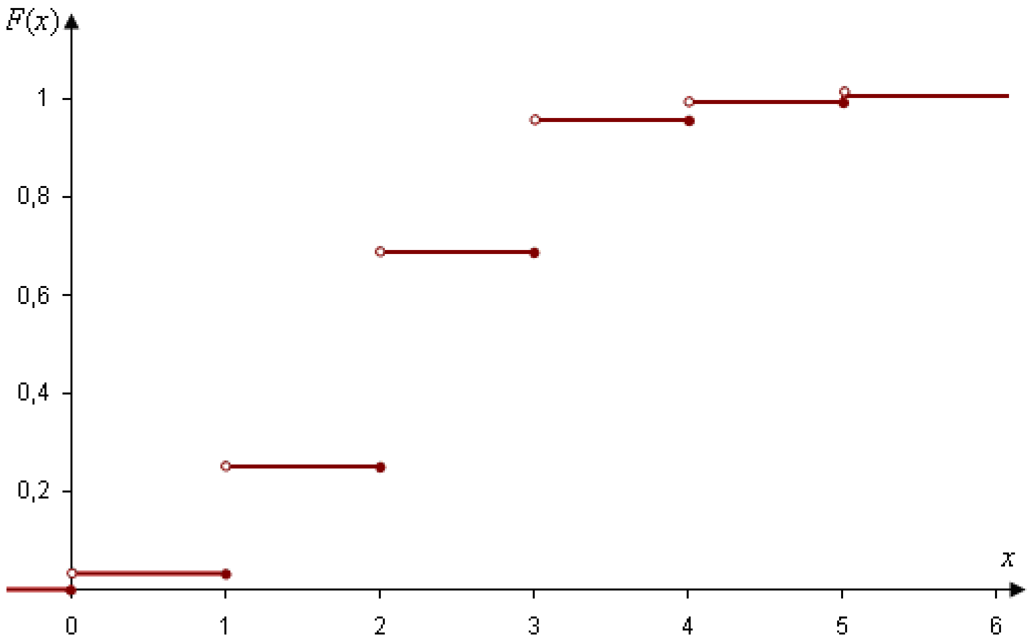
Histogram:



Tabulka pro distribuční funkci:

x_i	0	1	2	3	4	5	6
p_i	$\frac{21}{792}$	$\frac{175}{792}$	$\frac{350}{792}$	$\frac{210}{792}$	$\frac{35}{792}$	$\frac{1}{792}$	
$F(x_i)$	0	$\frac{21}{792}$	$\frac{196}{792}$	$\frac{546}{792}$	$\frac{756}{792}$	$\frac{791}{792}$	1

Graf:



3.3. Spojitá náhodná veličina

Také u spojité náhodné veličiny se užívá k jejímu popisu **distribuční funkce *F*(*x*)**, která je definovaná stejně jako u diskrétní náhodné veličiny vztahem:

$F(x_i) = P(X < x_i)$

Vlastnosti *F*(*x*) (společné pro spojitou i diskrétní náhodnou veličinu):

- $0 \leq F(x) \leq 1$
- $P(x_1 \leq X < x_2) = F(x_2) - F(x_1)$ pro $x_1 < x_2$
- $F(x)$ je neklesající funkce
- $F(-\infty) = 0$, $F(\infty) = 1$
- $F(x)$ je zleva spojitá v bodech $x = x_i$, $i = 1,2,...$, diskrétní náhodné veličiny a spojitá v ostatních bodech.

Druhou vlastnost je možné zapsat také: $P(x \leq X < x + h) = F(x + h) - F(x)$.
Pro $h \to 0$ levá strana $\to P(X = x)$ a pravá $\to 0$ (tedy $P(X = x) = 0$).
Proto nemá smysl definovat pro spojitou náhodnou veličinu pravděpodobnostní funkci

$p(x) = \mathrm{P}(X = x)$. Zavádíme tedy jinou funkci, která se nazývá **hustota pravděpodobnosti**:

Definice 3.3.1.

Hustota pravděpodobnosti náhodné veličiny *X* definované na intervalu $\langle a,b \rangle$ je nezáporná, reálná funkce definovaná vztahem:

$$f(x) = \lim_{h\to 0} \frac{P(x \leq X < x+h)}{h},$$

kde pro $x \notin \langle a,b \rangle$ je $f(x) = 0$; x, $x+h \in \langle a,b \rangle$

Vlastnosti $f(x)$ a $F(x)$ spojité náhodné veličiny

- pro $\forall x \in R$ platí: $f(x) \geq 0$
- $\int_a^b f(x)\,dx = 1$ (obecně $\int_{-\infty}^{\infty} f(x)\,dx = 1$); a, b jsou krajní meze intervalu, ve kterém je $f(x)$ různá od nuly)
- $f(x) = F'(x)$ ($F(x)$ je primitivní funkcí $f(x)$)
- $F(x) = P(X < x) = \int_a^x f(x)\,dx$ resp. $= \int_{-\infty}^x f(x)\,dx$
- $P(x_1 \leq X < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\,dx$

Řešené úlohy

Příklad 3.3.1. Náhodná veličina X je dána distribuční funkcí:

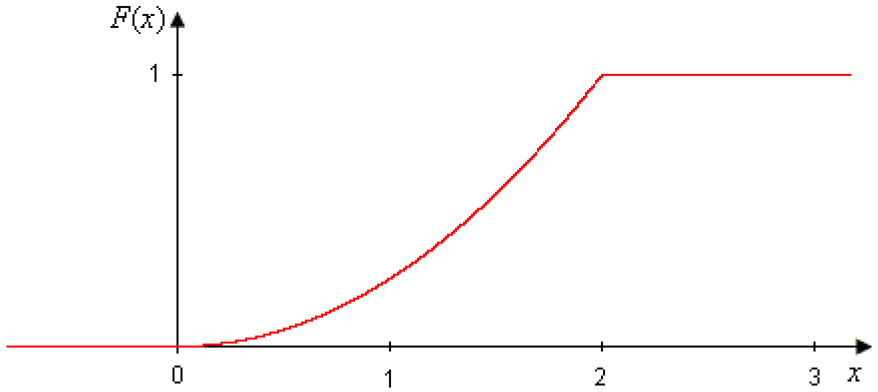
$$F(x) = \begin{cases} 0 & x \leq 0 \\ \dfrac{x^2}{4} & 0 < x \leq 2 \\ 1 & x > 2 \end{cases}$$

Určete $f(x)$, znázorněte graficky $F(x)$, $f(x)$, vypočtěte $P(0{,}4 \leq X < 1{,}6)$

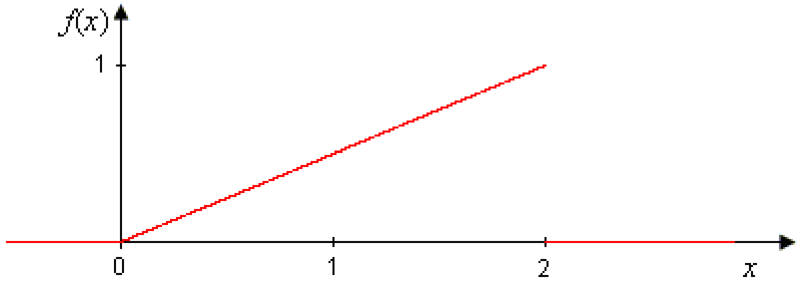
Řešení: Hustotu pravděpodobnosti získáme zderivováním distribuční funkce:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \dfrac{x}{2} & 0 < x \leq 2 \\ 0 & x > 2 \end{cases}$$

Graf distribuční funkce:



Graf hustoty pravděpodobnosti:



$P(0{,}4 \le X < 1{,}6) = F(1{,}6) - F(0{,}4) = 0{,}64 - 0{,}04 = 0{,}6$

Příklad 3.3.2. Hustota pravděpodobnosti náhodné veličiny X má tvar:

$$f(x) = \begin{cases} 0 & x < 0 \\ a.\sin x & 0 \le x < \pi \\ 0 & x \ge \pi \end{cases}$$

Určete koeficient a, distribuční funkci $F(x)$ a $P\left(\frac{\pi}{2} < X < 2\pi\right)$.

Řešení: Nejdříve určíme koeficient a:

$$\int_0^{\pi} a.\sin x dx = 1$$

$$a.\left[-\cos x\right]_0^{\pi} = 1$$

$$a.2 = 1$$

$$a = \frac{1}{2}$$

$F(x)$ je primitivní funkcí $f(x)$. Jestliže integrujeme $f(x)$, obdržíme:

$$F(x) = \begin{cases} C_1 & x < 0 \\ -\frac{1}{2}\cos x + C_2 & 0 \le x < \pi \\ C_3 & x \ge \pi \end{cases}$$

Hodnoty konstant C_1, C_3 zjistíme z okrajových podmínek distribuční funkce:

$F(-\infty) = 0$, $F(\infty) = 1$. Takže $C_1 = 0$, $C_3 = 1$.

Pro vypočtení konstanty C_2 využijeme spojitosti distribuční funkce. Víme, že:

$$F(0) = 0$$

$$-\tfrac{1}{2}\cos 0 + C_2 = 0$$

$$C_2 = \frac{1}{2}$$

Distribuční funkce má tedy tvar:

$$F(x)=\begin{cases}0 & x<0\\ -\frac{1}{2}\cos x+\frac{1}{2} & 0\le x<\pi\\ 1 & x\ge\pi\end{cases}$$

Výpočet hledané pravděpodobnosti:

$$P\left(\tfrac{\pi}{2}<X<2\pi\right)=F(2\pi)-F\left(\tfrac{\pi}{2}\right)=1-\left(-\tfrac{1}{2}\cos\tfrac{\pi}{2}+\tfrac{1}{2}\right)=\frac{1}{2}$$

Příklad 3.3.3. Určete konstanty A, B tak, aby funkce $F(x) = A + B.\arctan x$, definovaná pro všechna reálná čísla, byla distribuční funkcí rozložení náhodné veličiny.

Řešení:

$$F(-\infty)=0$$
$$F(\infty)=1$$

$$A+B.\arctan(-\infty)=0$$
$$A+B.\arctan(\infty)=1$$

$$A+B.\left(-\frac{\pi}{2}\right)=0$$
$$A+B.\left(\frac{\pi}{2}\right)=1$$

$$A=\frac{1}{2}$$
$$B=\frac{1}{\pi}$$

Poznámka

Rozdělení určené distribuční funkcí z předchozího příkladu se nazývá Cauchyho rozdělení náhodné veličiny.

Pro získání komplexnějšího pohledu na problematiku náhodné veličiny, doporučujeme, přečíst si Úvod do teorie informací. Zde se dozvíte více o pojmu neurčitosti.

3.4. Číselné charakteristiky náhodné veličiny

Náhodná veličina X je jednoznačně určena rozdělením pravděpodobnosti pomocí pravděpodobnostní funkce nebo distribuční funkce (popř. hustoty pravděpodobnosti). Tyto funkce jsou však často poměrně složité a jejich určení pracné. Proto je výhodné shrnout informace o náhodné veličině do několika čísel, které ji dostatečně charakterizují. Tato čísla nazýváme **číselné charakteristiky** a dělíme je:

a) podle způsobu konstrukce na charakteristiky:

- momentové
- kvantilové
- ostatní

b) podle toho, které vlastnosti rozdělení pravděpodobnosti charakterizují na charakteristiky:

- polohy
- variability
- šikmosti
- špičatosti

3.4.1. Momentové charakteristiky náhodné veličiny

Jsou konstruovány na základě počátečního momentu μ_k nebo centrálního momentu ν_k:

Definice 3.4.1.

Počáteční (obecný) moment *k*-tého stupně $\boldsymbol{\mu_k}$ náhodné veličiny X je střední hodnota k-té mocniny náhodné veličiny:

$$\mu_k = \begin{cases} \sum_i x_i^k . p(x_i) & \text{pro diskrétní náhodnou veličinu} \\ \int_{-\infty}^{\infty} x^k . f(x)\, dx & \text{pro spojitou náhodnou veličinu} \end{cases}$$

Centrální moment *k*-tého stupně ν_k náhodné veličiny *X* je:

$$\upsilon_k = \begin{cases} \sum_i (x_i-\mu)^k . p(x_i) & \text{pro diskrétní náhodnou veličinu} \\ \int_{-\infty}^{\infty} (x-\mu)^k . f(x)\, dx & \text{pro spojitou náhodnou veličinu} \end{cases},$$

kde $\mu = \mu_1$ je počáteční moment 1. stupně náhodné veličiny *X*.

Poznámka

Praktický význam mají čtyři momentové charakteristiky: μ_1, ν_2, ν_3, ν_4

První počáteční moment μ_1

představuje **střední hodnotu** náhodné veličiny *X*

Bývá označován: $\mu_1 = E(X) = \mu$

tedy:

$$E(X) = \mu = \begin{cases} \sum_i x_i . p(x_i) & \text{pro diskrétní náhodnou veličinu} \\ \int_{-\infty}^{\infty} x . f(x)\, dx & \text{pro spojitou náhodnou veličinu} \end{cases}$$

Pro střední hodnotu platí:

1. $E(c) = c$, kde *c* je konstanta
2. $E(c.X) = c.E(X)$
3. $E(X \pm Y) = E(X) \pm E(Y)$
4. $E(X.Y) = E(X).E(Y)$, jsou-li *X* a *Y* nezávislé

Druhý centrální moment ν_2
představuje **rozptyl** (disperzi, varianci)
Označujeme: $\nu_2 = D(X) = \sigma^2$

$$D(X) = \sigma^2 = \begin{cases} \sum_i (x_i-\mu)^2 . p(x_i) & \text{pro diskrétní náhodnou veličinu} \\ \int_{-\infty}^{\infty} (x-\mu)^2 . f(x)\, dx & \text{pro spojitou náhodnou veličinu} \end{cases}$$

Pro rozptyl platí:

1. $D(c) = 0$, kde c je konstanta
2. $D(c.X) = c^2.D(X)$
3. $D(X + Y) = D(X) + D(Y)$, jsou-li X a Y nezávislé
4. $\sqrt{D(X)} = \sqrt{\sigma^2} = \sigma$ . . . se nazývá **směrodatná odchylka**

Rozptyl a směrodatná odchylka charakterizují rozptýlenost hodnot náhodné veličiny X kolem střední hodnoty μ.

Další dvě číselné charakteristiky jsou vyjádřeny pomocí normovaných momentů.

Normovaný moment *r*-tého stupně $\overline{\nu_r}$ náhodné veličiny X je určen vztahem

$$\overline{\nu_r} = \frac{\nu_r}{\sigma^r},$$

v němž ν_r značí centrální moment r-tého stupně a σ^r je r-tá mocnina směrodatné odchylky náhodné veličiny X.

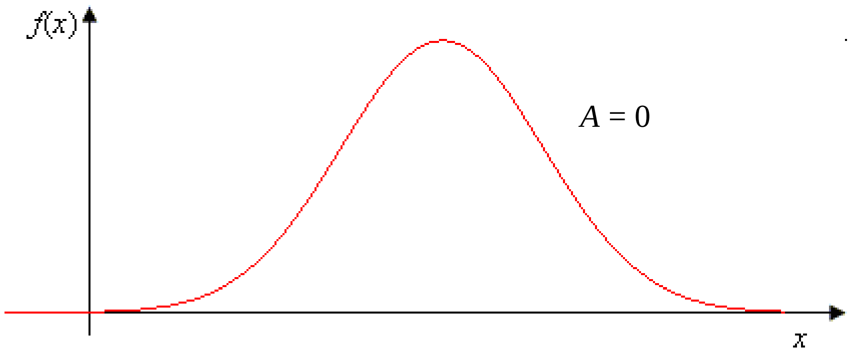
Třetí centrální moment ν_3

slouží k určení **koeficientu asymetrie**, který označujeme $\overline{\nu_3} = A$

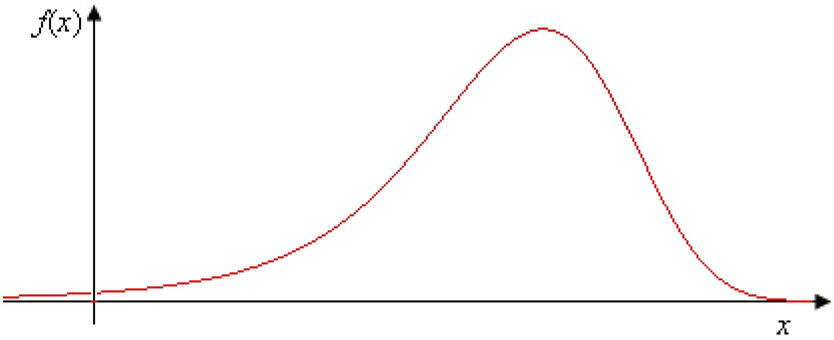
$A = \overline{\upsilon_3} = \dfrac{\upsilon_3}{\sigma^3}$, kde

$$\upsilon_3 = \begin{cases} \sum\limits_i (x_i - \mu)^3.p(x_i) & \text{pro diskrétní náhodnou veličinu} \\ \int\limits_{-\infty}^{\infty} (x - \mu)^3.f(x)\,dx & \text{pro spojitou náhodnou veličinu} \end{cases}$$

Vyjadřuje, do jaké míry a na kterou stranu je rozložení zešikmeno, nebo jestli je symetrické:



zešikmení vlevo: $A < 0$



zešikmení vpravo: $A > 0$

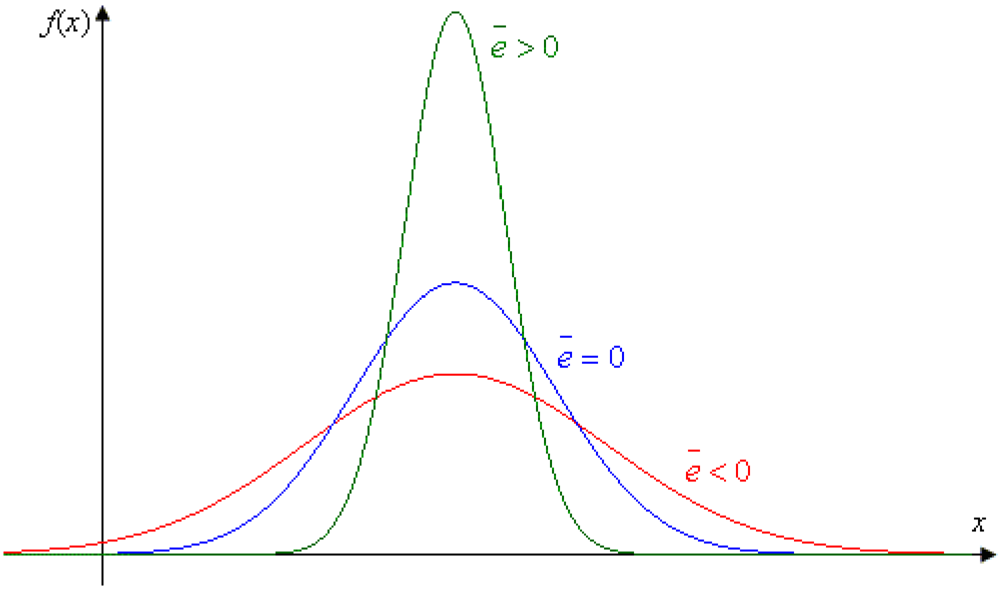
Čtvrtý centrální moment v_4

slouží k výpočtu **koeficientu špičatosti** (excesu), který značíme $\bar{e}$.

$\bar{e} = \overline{\upsilon_4} = \dfrac{\upsilon_4}{\sigma^4} - 3$, kde

$$\upsilon_4 = \begin{cases} \sum\limits_i (x_i - \mu)^4 . p(x_i) & \text{pro diskrétní náhodnou veličinu} \\ \int\limits_{-\infty}^{\infty} (x - \mu)^4 . f(x)\, dx & \text{pro spojitou náhodnou veličinu} \end{cases}$$

Informuje o koncentrovanosti hodnot dané veličiny kolem její střední hodnoty.



Výpočet centrálních momentů lze provádět podle výše uvedeného a nebo s využitím vztahů mezi μ_k a v_k:

- $v_2 = \mu_2 - \mu_1^2$
- $v_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$
- $v_4 = \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4$

  $\vdots$

- $$\upsilon_k = \binom{k}{0}\mu_k\mu_1^0 - \binom{k}{1}\mu_{k-1}\mu_1^1 + \binom{k}{2}\mu_{k-2}\mu_1^2 + \ldots + (-1)^k\binom{k}{k}\mu_1^k$$

Řešené úlohy

Příklad 3.4.1.

Náhodná veličina X je dána tabulkou. Určete její číselné charakteristiky

x_i	1	2	3	4
p_i	0,3	0,1	0,4	?

Řešení: $p_4 = 1 - (p_1 + p_2 + p_3) = 0{,}2$

$$E\;\;X\;\;=\mu=\sum_{i=1}^{4} x_i.p\;\;x_i\;\;=1.0{,}3+2.0{,}1+3.0{,}4+4.0{,}2=2{,}5$$

$$D\;\;X\;\;=\sigma^2=\sum_{i=1}^{4}\;\;x_i-\mu\;\;^2.p\;\;x_i\;\;=$$

$$=\;\;1-2{,}5\;\;^2.0{,}3+\;\;2-2{,}5\;\;^2.0{,}1+\;\;3-2{,}5\;\;^2.0{,}4+\;\;4-2{,}5\;\;^2.0{,}2=1{,}25$$

Další charakteristiky vypočteme pomocí následující tabulky:

x_i	1	2	3	4	Σ
p_i	0,3	0,1	0,4	0,2	-
$x_i.p(x_i)$	0,3	0,2	1,2	0,8	2,5
$x_i^2.p(x_i)$	0,3	0,4	3,6	3,2	7,5
$x_i^3.p(x_i)$	0,3	0,8	10,8	12,8	24,7
$x_i^4.p(x_i)$	0,3	1,6	32,4	51,2	85,5

Tedy:

$$A=\frac{\nu_3}{\sigma^3}=\frac{\mu_3-3\mu_1\mu_2+2\mu_1^3}{\sigma^3}=\frac{24{,}7-3.2{,}5.7{,}5+2.2{,}5^3}{\sqrt{1{,}25}\;\;^3}=-0{,}21$$

$$\bar{e}=\frac{\nu_4}{\sigma^4}-3=\frac{\mu_4-4\mu_3\mu_1+6\mu_2\mu_1^2-3\mu_1^4}{\sigma^4}=\ldots=-1{,}36$$

Příklad 3.4.2. Náhodná veličina X má hustotu pravděpodobnosti:

$$f\;\;x\;\;=\begin{cases}2x & \text{pro } x\in\langle 0,1\rangle\\ 0 & \text{pro ostatní } x\end{cases}$$

Určete její číselné charakteristiky

Řešení:

$$E\;\;X\;\;=\mu=\int_0^1 x.2xdx=\left[\frac{2x^3}{3}\right]_0^1=\frac{2}{3}=0{,}\bar{6}$$

$$\mu_2=\int_0^1 x^2.2xdx=\left[\frac{x^4}{2}\right]_0^1=\frac{1}{2}=0{,}5$$

$$\mu_3 = \int_0^1 x^3.2x dx = \left[\frac{2x^5}{5}\right]_0^1 = \frac{2}{5} = 0,4$$

$$\mu_4 = \int_0^1 x^4.2x dx = \left[\frac{x^6}{3}\right]_0^1 = \frac{1}{3} = 0,\overline{3}$$

$$D\ \ X\ = \mu_2 - \mu_1^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18} = 0,0\overline{5}$$

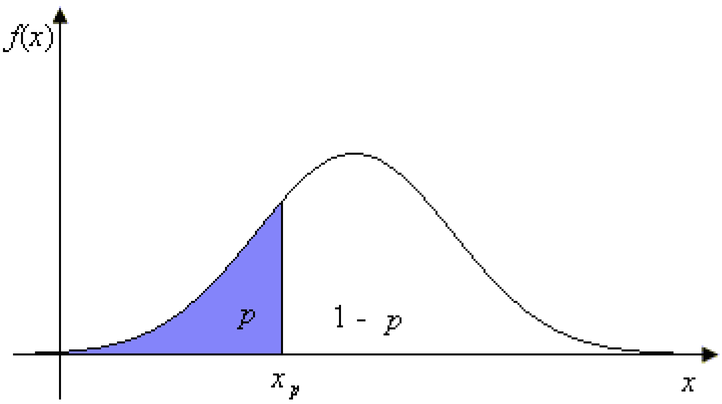
$$A = \frac{\nu_3}{\sigma^3} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{\sigma^3} = \ldots = -0,4\overline{3}$$

$$\bar{e} = \frac{\nu_4}{\sigma^4} - 3 = \frac{\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4}{\sigma^4} = \ldots = -0,4$$

3.4.2. Kvantilové charakteristiky náhodné veličiny

- jsou obvykle odvozeny pomocí distribuční funkce $F(x)$
- jsou určovány pro spojitou náhodnou veličinu, pro diskrétní náhodnou veličinu nebývá jejich určení jednoznačné

Definice 3.4.2.
Nechť $F(x)$ je distribuční funkce spojité náhodné veličiny X. Pak hodnota x_p, pro kterou platí $F(x_p) = p$, kde $p \in \langle 0,1 \rangle$, se nazývá ***p*-kvantil**



p-kvantil dělí plochu pod grafem hustoty pravděpodobnosti v poměru p:(1-p)

Nejužívanější kvantily:

- kvartily: $x_{0,25}$, $x_{0,50}$, $x_{0,75}$
  - rozdělí obor možných hodnot na čtyři části, v nichž se náhodná veličina nachází s pravděpodobností 0,25

- decily: $x_{0,1}$, $x_{0,2}$, ..., $x_{0,9}$
  - rozdělí obor možných hodnot na deset částí se stejnou pravděpodobností výskytu

- percentily: $x_{0,01}$, $x_{0,02}$, ..., $x_{0,99}$
  - rozdělí obor možných hodnot na sto částí se stejnou pravděpodobností výskytu

$x_{0,5} = Me$ . . . **medián**: dělí plochu pod křivkou hustoty pravděpodobnosti na dvě stejné části

Řešené úlohy

Příklad 3.4.3. Určete první decil $x_{0,1}$ a třetí kvartil $x_{0,75}$ pro

$$f(x) = \begin{cases} \frac{1}{2} & \text{pro } x \in \langle 0,2 \rangle \\ 0 & \text{pro ostatní } x \end{cases}$$

Řešení:

$$F(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 0) \\ \frac{x}{2} & \text{pro } x \in \langle 0,2 \rangle \\ 1 & \text{pro } x \in (2, \infty) \end{cases}$$

$$F(x_{0,1}) = 0,1 \quad F(x_{0,75}) = 0,75$$

$$\frac{1}{2} x_{0,1} = 0,1 \qquad \frac{1}{2} x_{0,75} = 0,75$$

$$x_{0,1} = 0,2 \qquad x_{0,75} = 1,5$$

Modus: *Mo* - je hodnota, v níž nabývá frekvenční funkce maxima:

- u diskrétní náhodné veličiny je to hodnota, v níž pravděpodobnostní funkce $p(x_i)$ dosahuje maxima

- u spojité náhodné veličiny je to hodnota, v níž hustota pravděpodobnosti $f(x)$ nabývá lokálního maxima

Řešené úlohy

Příklad 3.4.4. Náhodná veličina X má hustotu pravděpodobnosti:

$$f(x) = \begin{cases} \frac{1}{2}x^2 e^{-x} & \text{pro } x \in (0,\infty) \\ 0 & \text{pro } x \notin (0,\infty) \end{cases}.$$

Určete modus.

Řešení: Modus je hodnota, v níž frekvenční funkce (v našem případě hustota pravděpodobnosti) nabývá maxima. Maximum funkce vypočteme pomocí první derivace:

$$f'(x) = x.e^{-x} - \tfrac{1}{2}x^2.e^{-x}$$

První derivace položíme rovnu nule:

$$x.e^{-x}\left(1-\tfrac{1}{2}x\right) = 0$$

Tato rovnice má dvě řešení:

$x = 0$ ... toto řešení není přípustné, nula neleží v definičním oboru

$x = 2$ ... lehce ověříme, že se skutečně jedná o maximum

$Mo = 2$

3.4.3. Shrnutí

- **Charakteristiky polohy**
  $E(X)$, Me, Mo, kvantily. Určují jakýsi "střed", kolem něhož kolísají hodnoty náhodné veličiny X.

- **Charakteristiky variability**
  $D(X)$, σ, ... Ukazují rozptýlenost hodnot náhodné veličiny kolem střední hodnoty

- **Charakteristiky šikmosti a špičatosti**
  Charakterizují průběh rozdělení náhodné veličiny X

Úlohy k samostatnému řešení

Náhodná veličina

3.1. Třikrát vystřelíme na cíl. Pravděpodobnost zásahu při každém výstřelu je $p = 0{,}7$.

Určete:

a) pravděpodobnostní funkci počtu zásahů při třech nezávislých výsledcích,

b) distribuční funkci a její graf.

3.2. Házeme třikrát kostkou. Nechť náhodná veličina X znamená počet padnutí šestky.

Určete:

a) pravděpodobnostní funkci a její graf,

b) sestrojte graf distribuční funkce.

3.3. Náhodná veličina X je dána distribuční funkcí:

$$F(x) = \begin{cases} 0 & \text{pro } x < 3 \\ \dfrac{x}{3} - 1 & \text{pro } 3 \le x < 6 \\ 1 & \text{pro } x \ge 6 \end{cases}$$

Určete $f(x)$, znázorněte graficky $f(x)$, $F(x)$ a $P(1{,}5 \le X \le 4)$.

3.4. Hustota pravděpodobnosti náhodné veličiny X má tvar:

$$f(x) = \begin{cases} 0 & \text{pro } x < 1 \\ x - \dfrac{1}{2} & \text{pro } 1 \le x < 2 \\ 0 & \text{pro } x \ge 2 \end{cases}$$

Určete distribuční funkci

3.5. Hustota pravděpodobnosti náhodné veličiny X má tvar:

$$f(x) = \begin{cases} 0 & \text{pro } x < 0 \\ cx(1-x) & \text{pro } 0 \le x < 1 \\ 0 & \text{pro } x \ge 1 \end{cases}$$

Určete koeficient c, distribuční funkci $F(x)$ a $P(X > 0{,}2)$.

3.6. Distribuční funkce náhodné veličiny X má tvar:

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x \quad \text{pro } -\infty < x < \infty.$$

Určete pravděpodobnost, že náhodná veličina X nabývá hodnot z intervalu $(0,1)$.

3.7. Dva hráči hrají společenskou hru. Pravděpodobnost výhry hráče A je 2/3, hráče B 1/3. Hráči opakují hru tolikrát, až vyhraje hráč A. Určete zákon rozložení náhodné veličiny, která značí počet uskutečněných her.

3.8. Určete zákon rozložení náhodné veličiny, která značí součet ok při hodu

a) jednou kostkou,

b) dvěma kostkami,

c) třemi kostkami.

3.9. Střelec střílí 10-krát na cíl. Za každý zásah získává 3 body, nezasáhne-li, ztrácí 1 bod. Pravděpodobnost zásahu při jednom výstřelu daného střelce je 2/3. Určete zákon rozložení počtu bodů, které střelec může získat.

3.10. Pokus spočívá ve třech nezávislých hodech mincí. Pro náhodnou veličinu značící počet padnutí líců sestrojte funkci rozložení.

3.11. Hrací kostkou házíme n-krát. Najít funkce rozložení počtu padnuvších šestek.

3.12. Dokažte, že pro n = 1,2, …je výraz

$$p_n = \frac{1}{n} - \frac{1}{n+1}$$

zákonem rozložení diskrétní náhodné veličiny. Určete pravděpodobnosti $P(X < 3)$,

$P \quad X \leq 10$.

3.13. Výsledkem určitého pokusu je celé kladné číslo n s pravděpodobností nepřímo úměrnou n^2. Určete zákon rozložení náhodné veličiny.

3.14. Je dána funkce rozložení:

$$F(x) = \begin{cases} 0 & \text{pro } x < 1 \\ x-1 & \text{pro } 1 \leq x < 2 \\ 1 & \text{pro } x \geq 2 \end{cases}.$$

Určete k této funkci

a) hustotu rozložení $f(x)$,

b) pravděpodobnost $P\left(\frac{6}{5} \leq X < \frac{3}{2}\right)$.

3.15. Určete,

a) pro jaká *A*, *B* bude $F \quad x \ = A + \dfrac{B}{1+x^2}$ funkcí rozložení náhodné proměnné pro

$x \in \ 0, \infty \ $,

b) příslušnou hustotu rozložení.

3.16. Určete,

a) pro jaké *C* bude funkce $F \quad x \ = \sin Cx$ funkcí rozložení náhodné proměnné

pro $x \in \left\langle 0, 2\pi \right\rangle$,

b) příslušnou hustotu rozložení,

c) pravděpodobnost $P\left(\dfrac{\pi}{2} \le X < \dfrac{3\pi}{2} \right)$.

3.17. Určete

a) konstanty *A*, *B* tak, aby funkce $F \quad x \ = A + B.e^{-x}$ byla funkcí rozložení náhodné

veličiny pro $x \in \ 0, \infty \ $,

b) pravděpodobnost $P \ 1 \le X < 4 \ $,

c) hustotu rozložení *f(x)*.

3.18. Která z uvedených funkcí je pravděpodobnostní funkcí náhodné veličiny X , která nabývá hodnot 0, 2, 4, 6:

a) $f \quad x \ = \dfrac{1}{x}$

b) $f \quad x \ = \dfrac{c}{x+1}$

c) $f \quad x \ = \dfrac{x^2 - 4}{2}$

3.19. Náhodná veličina *X* je určena tabulkou:

X	-2	0	2	4	6
p	0,1	?	0,2	0,3	0,2

Určete hodnotu pravděpodobnosti pro *X* = 0, distribuční funkci a pravděpodobnost jevu, že náhodná veličina nabude kladných hodnot.

3.20. Cauchyho rozdělení náhodné veličiny *X* definované pro všechna reálná čísla má distribuční funkci $F(x) = a + b.\arctan x$. Určete konstanty *a, b*, hustotu pravděpodobnosti a pravděpodobnost, že *X* leží v intervalu $\left\langle \frac{\sqrt{3}}{3};1 \right)$.

3.21. Distribuční funkce Rayleighova rozdělení spojité náhodné veličiny má tvar:

$F(x) = C - e^{-\frac{x^2}{2\sigma^2}}$, $x > 0$. Určete konstantu *C* a hustotu pravděpodobnosti *f(x)*.

3.22. Distribuční funkce arkussinového rozložení pravděpodobnosti má tvar:

$$F(x) = \begin{cases} 0 & \text{pro } x < -1 \\ a + b.\arcsin x & \text{pro -1} \le x \le 1 \\ 1 & \text{pro } x > 1 \end{cases}$$

Určete konstanty *a, b* a hustotu pravděpodobnosti *f(x)*.

3.23. Je funkce $F(x) = \sin x$ distribuční funkcí náhodné veličiny *X* v intervalu

a) $\langle 0, \pi \rangle$,

b) $\left\langle 0, \frac{\pi}{2} \right\rangle$?

3.24. Náhodná veličina *X* je určena distribuční funkcí:

$$F(x) = \begin{cases} 0 & \text{pro } x < 2 \\ 2x - 4 & \text{pro } x \in \langle 2;2{,}5 \rangle \\ 1 & \text{pro } x > 2{,}5 \end{cases}.$$

Vypočítejte hustotu pravděpodobnosti náhodné veličiny *X*, pravděpodobnost toho, že *X* je menší než 7 / 3 a nakreslete grafy pravděpodobnostní a distribuční funkce.

3.25. Hustota pravděpodobnosti náhodné veličiny má tvar:

$$f(x) = \begin{cases} 0 & \text{pro } x < 0 \\ C.x.e^{-x} & \text{pro } x \ge 0 \end{cases}$$

Určete konstantu *C*, $P(0 \le X < 2)$ a distribuční funkci.

Číselné charakteristiky náhodné veličiny

3.26. Náhodná veličina X je dána tabulkou rozdělení pravděpodobnosti:

x_i	0	1	2	3
p_i	0,1	0,2	0,3	0,4

Určete střední hodnotu, rozptyl, koeficient asymetrie a špičatosti.

3.27. Pravděpodobnost zásahu cíle při každém ze čtyř výstřelů je 0,8. Nechť náhodná veličina X představuje počet zásahů cíle.

a) určete rozdělení pravděpodobnosti náhodné veličiny

b) vypočtěte její střední hodnotu, disperzi a směrodatnou odchylku

3.28. V městě byl po dobu 60 dnů evidován počet dopravních nehod v průběhu každého dne a podle počtu nehod v jednom dni vytvořena následující tabulka:

počet nehod / den	0	1	2	3	4	5	6
počet dnů s uvedeným počtem nehod	4	28	10	7	6	4	1

Pro počet nehod v jednom dni jako náhodnou proměnnou sestrojit zákon rozložení, střední hodnotu a disperzi.

(řešení v excelu)

3.29. Výsledkem náhodného pokusu je náhodná veličina nabývající hodnot $1/n$ (n je přirozené číslo) s pravděpodobnostmi nepřímo úměrnými 3^n. Určit střední hodnotu této náhodné veličiny.

(řešení v excelu) (jiná realizace řešení v excelu)

3.30. Náhodná veličina X má hustotu pravděpodobnosti:

$$f(x)=\begin{cases} 3x^2 & \text{pro } x\in\langle 0,1\rangle \\ 0 & \text{pro } x\notin\langle 0,1\rangle \end{cases}$$

Určete $E(x)$, $D(x)$

3.31. Náhodná veličina X má hustotu pravděpodobnosti:

$$f(x)=\begin{cases} \dfrac{3}{x^4} & \text{pro } x\in\langle 1,\infty) \\ 0 & \text{pro } x\notin\langle 1,\infty) \end{cases}$$

Určete $F(x)$, $E(x)$, $D(x)$, směrodatnou odchylku.

3.32. Určete střední hodnotu a rozptyl náhodné veličiny *X*, jejíž distribuční funkce má tvar:

$$F(x)=\begin{cases} 0 & \text{pro } x<0 \\ \dfrac{x}{2\pi} & \text{pro } x\in\langle 0,2\pi\rangle \\ 1 & \text{pro } x>2\pi \end{cases}$$

3.33. Házeme dvěma hracími kostkami. Určete rozdělení pravděpodobnosti součtu hozených bodů a modus.

3.34. Házeme třikrát mincí. Náhodná veličina *X* znamená hození líce. Určete rozdělení pravděpodobnosti a modus.

3.35. Náhodná veličina *X* má hustotu pravděpodobnosti:

$$f(x)=\begin{cases} \dfrac{1}{2}x^2 e^{-x} & \text{pro } x\in\langle 0,\infty) \\ 0 & \text{pro } x\notin\langle 0,\infty) \end{cases}$$. Určete modus.

3.36. Náhodná veličina *X* má hustotu pravděpodobnosti:

$$f(x)=\begin{cases} 2x & \text{pro } x\in\langle 0,1\rangle \\ 0 & \text{pro } x\notin\langle 0,1\rangle \end{cases}$$. Určete kvartily.

3.37. Náhodná veličina *X* má distribuční funkci:

$$F(x)=\begin{cases} 0 & \text{pro } x<2 \\ 2x-4 & \text{pro } x\in\langle 2;2{,}5\rangle \\ 1 & \text{pro } x>2{,}5 \end{cases}$$. Určete první tři decily.

3.38. Funkce $f(x)=C(2x-x^2)$ má být hustotou rozložení pravděpodobnosti pro $x\in\langle 0,2\rangle$.

Určete

a) konstantu *C*,

b) funkci rozložení *F(x)*,

c) střední hodnotu příslušné náhodné veličiny,

d) disperzi a směrodatnou odchylku,

e) pravděpodobnost *P(X<1)*.

3.39. Funkce $f(x)=Ax\sin x$ je funkcí hustoty rozložení pravděpodobnosti pro $x\in\langle 0,\pi\rangle$.

Určete

a) konstantu *A*

b) funkci *F(x)*,

c) střední hodnotu *E(X)*

d) disperzi *D(X)*

3.40. Funkce rozložení náhodné veličiny X má tvar

$$F(x)=\begin{cases}0 & \text{pro } x<-1\\ A+B.\arcsin x & \text{pro -}1\le x<1\\ 1 & \text{pro } x\ge 1\end{cases}$$. Určete

a) konstanty *A, B*

b) hustotu rozložení *f(x)*

c) střední hodnotu *E(X)*

d) disperzi *D(X)*

3.41. Určete střední hodnotu a rozptyl náhodné veličiny, která má hustotu rozložení ve tvaru

$f\quad x\ =\frac{1}{2}.e^{-|x|}$ (Laplaceovo rozložení).

3.42. Trolejbusy městské dopravy odjíždějí ze stanice v pětiminutových intervalech. Cestující přišel ke stanici v libovolný okamžik. Určete střední hodnotu a disperzi doby jeho čekání na odjezd ze stanice.

3.43. Mějme náhodnou veličinu X , jejíž hustota rozložení je dána

funkcí $f\quad x\ = A.\cos kx,\ x\in\left\langle -\frac{\pi}{2k},\frac{\pi}{2k}\right\rangle,\ k>0$

Určete konstantu A, střední hodnotu a disperzi.

Výsledky úloh k samostatnému řešení

3.1. $$p\ \ x\ = \binom{3}{x}.0,7^{x}.0,3^{3-x}$$

3.2. $$p\ \ x\ = \binom{3}{x}\cdot\left(\frac{1}{6}\right)^{x}\cdot\left(\frac{5}{6}\right)^{3-x}$$

3.3. $$f(x)=\begin{cases}\frac{1}{3} & \text{pro } 3\leq x<6\\ 0 & \text{jinde}\end{cases}$$

$$P\ \ 1,5\ \ \leq X\ \leq 4=\tfrac{1}{3}$$

3.4. $$F(x)=\begin{cases}0 & \text{pro } x<1\\ \dfrac{x-1}{2} & \text{pro } x\in\langle 1,2\\ 1 & \text{pro } x\geq 2\end{cases}$$

3.5. $c = 6$

$$F(x)=\begin{cases}0 & \text{pro } x<0\\ 3x^2-2x^3 & \text{pro } x\in\langle 0,1\\ 1 & \text{pro } x\geq 1\end{cases}$$

$P(X > 0,2) = 0,896$

3.6. $\dfrac{\pi}{4}$

3.7. $p_k = 2 / 3^k$

3.8. *a*) $6.p_k = (1, 1, 1, 1, 1, 1)$

b) $36.p_k = (1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1)$

c) $216.p_k = (1,3,6,10,15,21,25,27,27,25,21,15,10,6,3,1)$

3.9.

x_k	-10	-6	-2	2	6	10	14	18	22	26	30
$3^{-10}.p_k$	1	20	180	960	3360	8064	13440	15360	11520	5120	1024

3.10. $p_k = C_k(3).\ 1 / 2^3$

3.11. $p_k = 1 / 6^{\mathrm{n}}.C_k(n).5^{n-k}$, $k = 0,\dots,n$

3.12. $P(X<3) = 2 / 3$; $P(X<=10) = 10 / 11$

3.13. $f(n)=\dfrac{6}{\pi^2}\cdot\dfrac{1}{n^2}$

3.14.
$$a)\ f(x)=\begin{cases}0 & \text{pro } x<1\\ 1 & \text{pro } x\in\langle 1,2)\\ 0 & \text{pro } x\geq 2\end{cases}$$

$b)\ \dfrac{3}{10}$

3.15. $A=1,\ B=-1,\ f(x)=\dfrac{2x}{\left(1+x^2\right)^2}$

3.16. $a)\ C=\dfrac{1}{4}$

$$b)\ f(x)=\begin{cases}0 & \text{pro } x<0\\ \dfrac{1}{4}\cos\dfrac{x}{4} & \text{pro } x\in\langle 0,2\pi\rangle\\ 0 & \text{pro } x>2\pi\end{cases}$$

$c)\ \sqrt{\dfrac{2-\sqrt{2}}{2}}=0,5412$

3.17. $a)\ A=1,\ B=-1$

$b)\ P(1\leq X<4)=\dfrac{e^3-1}{e^4}$

$c)\ f(x)=e^{-x},\ x\in(0,\infty)$

3.18. pouze *b*) pro *c* = 35 / 92

3.19. $P(X=0)=0,2,\ P(X>7)=0,7$

3.20. $a=\dfrac{1}{2},\ b=\dfrac{1}{\pi},\ f(x)=\dfrac{1}{\pi}\cdot\dfrac{1}{1+x^2},\ p=\dfrac{1}{12}$

3.21. $C=1,\ f(x)=\dfrac{x}{\sigma^2}.e^{\frac{-x^2}{2\sigma^2}}$

3.22.
$$a=\frac{1}{2},\ b=\frac{1}{\pi},\ f(x)=\begin{cases}\dfrac{1}{\pi}\cdot\dfrac{1}{\sqrt{1-x^2}} & -1\leq x\leq 1\\ 0 & \text{jinde}\end{cases}$$

3.23. pouze *b*)

3.24. $f(x)=\begin{cases}2 & \text{pro } 2\le x\le 2{,}5\\ 0 & \text{jinde}\end{cases}$, $P\left(X<\frac{7}{3}\right)=\frac{2}{3}$

3.25. $C=1,\ P\ \ 0\le X<2\ \ =1-3e^{-2},\ F(x)=\begin{cases}1-e^{-x} & \text{pro } x\ge 0\\ 0 & \text{jinde}\end{cases}$

3.26. 2; 1; -0,6; -0,8

3.27. *a*) $\binom{4}{x}.0{,}8^{x}.0{,}2^{4-x}$

b) 3,2; 0,64

3.30. 0,75; 0,0375

3.31. $E(x) = 1{,}5$; $D(x) = 0{,}75$

3.32. $E(x) = \pi$, $D(x) = \frac{\pi^2}{3}$

3.33. $Mo(x) = 7$

3.34. $p\ \ x\ \ =\binom{3}{x}.0{,}5^{x}.0{,}5^{3-x},\ x=0,1,2,3;\ Mo\ \ x\ \ =1,2$

3.35. $Mo(x) = 2$

3.36. $x_{0,25} = 0{,}5$

$x_{0,25}=\frac{\sqrt{2}}{2}$

$x_{0,75}=\frac{\sqrt{3}}{2}$

3.37. $x_{0,1} = 2{,}05$; $x_{0,2} = 2{,}1$; $x_{0,3} = 2{,}15$

3.38. C = 3 / 4 , $F(x) = 3/4\,(x^2 - x^3/3)$, $x_{stř} = 1$, $D(X) = 1/5$, $\sigma = \sqrt{(1/5)} = 0{,}4472$, $p = 1/$
2

3.39. $A = 1/\pi$, $F(x) = 1/\pi(\sin(x)-x\cos(x))$, $E(X) = \pi - 4/\pi$, $D(X) = 2 - 16/\pi^2$

3.40. $A = 1/2$, $B = 1/\pi$, $f(x) = 1/\pi\sqrt{(1-x^2)}$, $E(X) = 0$, $D(X) = 1/2$, $M_3 = 0$, $M_4 = 3/8$

3.41. $x_{stř} = 0$, $\sigma^2 = 2$

3.42. $f(x) = 1 / 5$, x in <0, 5> , $x_{stř} = 5 / 2(\text{min}) = 150(\text{s})$, $D = 25 / 12(\text{min}^2)$

3.43. $A = k / 2$, $E(X) = 0$, $D(X) = (\pi - 8) / 4\,\text{k}^2 \approx 0{,}4672 / k^2$

4. ZÁKLADNÍ TYPY ROZDĚLENÍ PRAVDĚPODOBNOSTI DISKRÉTNÍ NÁHODNÉ VELIČINY

Průvodce studiem

V této kapitole se seznámíte se základními typy rozložení diskrétní náhodné veličiny. Vašim úkolem by neměla být pouze základní pasivní znalost a orientace v rozloženích, ale měli byste se také naučit tato rozložení od sebe rozlišovat a bezpečně je rozpoznávat.

Předpokládané znalosti

Pojmy z kombinatoriky, pravděpodobnosti.

Cíle

Cílem této kapitoly je seznámení se základními typy rozložení diskrétní náhodné veličiny, odvození jejich základních číselných charakteristik.

Výklad

4.1. Alternativní rozdělení $A(p)$

Některé náhodné pokusy mohou mít pouze dva různé výsledky:

- pokus je úspěšný
- pokus je neúspěšný

Příslušná náhodná veličina X se pak nazývá **alternativní** (dvoubodová, nulajedničková).

Tato náhodná veličina nabývá tedy pouze dvou hodnot: 1 - v případě příznivého výsledku pokusu (jev A), 0 - v případě nepříznivého výsledku pokusu (jev $\overline{A}$).

Obor hodnot tedy obsahuje dva prvky $M = \{0,1\}$.

Používáme označení: $P(A) = P(X = 1) = p$

$$P(\overline{A}) = P(X = 0) = 1 - p$$

Definice 4.1.1.

Náhodná veličina X s pravděpodobnostní funkcí $P(X = 0) = 1 - p$, $P(X = 1) = p$ $(0 < p < 1)$ má **alternativní rozdělení** pravděpodobnosti $A(p)$ s parametrem p.

Řešené úlohy

Příklad 4.1.1. Hod mincí: $\Omega = \{\text{líc,rub}\}$

Jedná se o alternativní rozdělení $A\left(\frac{1}{2}\right)$.

Tedy: $M = \{0,1\}$; $X = \{0 \vee 1\}$

$$p(0) = \frac{1}{2}$$

$$p(1) = 1 - \frac{1}{2} = \frac{1}{2}$$

4.2. Rovnoměrné rozdělení *R*(*n*)

Definice 4.2.1.

Náhodná veličina X má **rovnoměrné rozdělení** $R(n)$ právě tehdy, když je pravděpodobnostní funkce určena vztahem:

$p(x) = \dfrac{1}{n}$, kde n je počet možných výsledků.

Řešené úlohy

Příklad 4.2.1. Hod kostkou: $M = \{1, 2, 3, 4, 5, 6\}$ - každý výsledek je stejně pravděpodobný.

Jedná se tedy o rovnoměrné rozdělení $R(6)$, $p(x) = \dfrac{1}{6}$

4.3. Binomické rozdělení *Bi*(*n*, *p*)

- popisuje četnost náhodného jevu v n nezávislých pokusech, v nichž má jev stále stejnou pravděpodobnost

Definice 4.3.1.

Náhodná veličina *X* má **binomické rozdělení** *Bi*(*n*, *p*) právě tehdy, když je pravděpodobnostní funkce určena vztahem:

$p(x) = \binom{n}{x}.p^x.(1-p)^{n-x}$, kde $x = 0, 1, \ldots, n$; *n* je počet pokusů a *p* je pravděpodobnost úspěšnosti v každém pokusu.

Binomické rozdělení je tedy příkladem diskrétního rozdělení pravděpodobnosti náhodné proměnné *X*, která může nabývat pouze $n+1$ hodnot. Při matematickém sestrojení binomického rozdělení vycházíme z Bernoulliova pokusu, který spočívá v tom, že v daném náhodném pokusu mohou nastat pouze dva stavy: *A*, $\overline{A}$ s pravděpodobností *p*, 1 - *p*. To lze modelovat tzv. binární náhodnou proměnnou *Y*, pro kterou platí: $P(Y = 1) = p$ a $P(Y = 0) = 1 - p$. Platí:

$E(Y) = 1.p + 0.(1 - p) = p$

$D(Y) = E(Y - p)^2 = p.(1 - p)^2 + (1 - p).p^2 = (1 - p).p$

Náhodná proměnná *X* vznikne jako součet *n* nezávislých binárních proměnných Y_i s hodnotami 0 nebo 1, které mají všechny stejné rozdělení určené parametrem *p*:

$$X = \sum_{i=1}^{n} Y_i$$

Z toho plyne:

Vlastnosti binomického rozdělení:

$E(X) = n.p$

$D(X) = n.p.(1 - p)$

Poznámka

Alternativní rozdělení A(p) *je vlastně speciálním případem binomického rozdělení pro* n = 1 (*A*(*p*) ~ *Bi*(1,*p*)).

Řešené úlohy

Příklad 4.3.1. Student VŠB Pepe má potíže s ranním vstáváním. Proto někdy zaspí a nestihne přednášku, která začíná již v 9 hodin. Pravděpodobnost, že zaspí, je 0,3. V semestru je 12 přednášek - tzn. 12 nezávislých pokusů dorazit na přednášku včas. Nalezněte pravděpodobnost, že Pepe nestihne přednášku v důsledku zaspání v polovině nebo více případů.

Řešení: Hledaná pravděpodobnost má hodnotu:

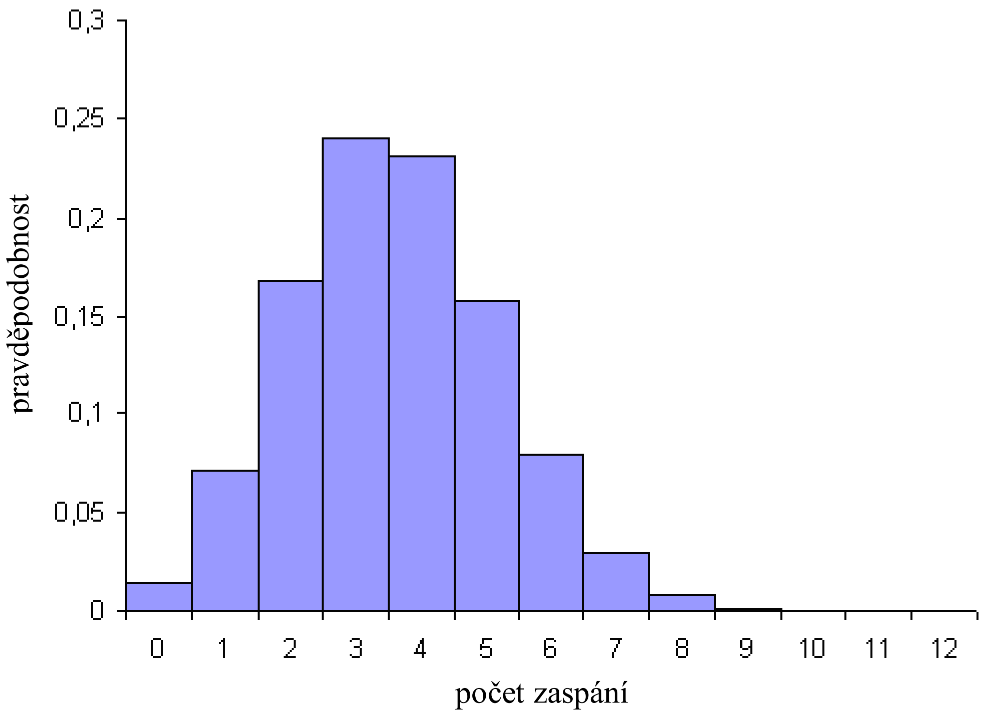
$$P(X \geq 6) = P(6) + P(7) + P(8) + P(9) + P(10) + P(11) + P(12) =$$

$$= \sum_{k=6}^{12} \binom{12}{x} .0,3^{k}.0,7^{12-k} \approx 0,118$$

Ruční výpočet by v tomto případě byl poměrně zdlouhavý. Máme-li ale k dispozici např. tabulkový procesor Excel, můžeme příklad snadno vypočíst pomocí distribuční funkce binomického rozdělení - v Excelu ji najdeme pod názvem BINOMDIST:

$P(X \geq 6) = 1 - P(X < 6) = 1 - F(6) = 1 -$ BINOMDIST(5;12;0,3;1) = 0,118

Rozdělení pravděpodobnosti pro tento příklad je znázorněno graficky na následujícím obrázku:



4.4. Poissonovo rozdělení *Po*(λ)

Toto rozdělení pravděpodobnosti, pojmenované podle francouzského matematika S. D. Poissona, mají náhodné proměnné, které popisují četnosti jevů s těmito vlastnostmi:

- to, že jev v daném intervalu (časovém, prostorovém) nastane (nenastane), nezávisí na tom, co se stalo jindy nebo jinde
- pro každý časový okamžik je pravděpodobnost jevu v malém časovém intervalu stejná (totéž platí v prostoru)
- neexistuje případ, že by nastaly dva jevy přesně v jednom časovém okamžiku nebo místě v prostoru

Průměrný počet výskytů zkoumaného jevu v daném úseku jednotkové délky označujeme λ.

Definice 4.4.1.

Náhodná veličina *X* má **Poissonovo rozdělení** *Po*(λ) právě tehdy, když má pravděpodobnostní funkce tvar:

$p(x) = \frac{\lambda^x}{x!}.e^{-\lambda}$ v daném jednotkovém úseku, kde $x = 0,1,2,\dots$; $\lambda > 0$ je parametr.

Případně $p(x) = \frac{(l\lambda)^x}{x!}.e^{-l\lambda}$ v úseku délky *l* (v *l*-násobku délky jednotkového úseku)

Pro charakteristiky Poissonova rozdělení platí:

- $E(x) = \lambda$
- $D(x) = \lambda$
- $A = \frac{1}{\sqrt{\lambda}}$
- $\bar{e} = \frac{1}{\lambda}$

Poznámka

S rostoucí hodnotou λ se toto rozdělení blíží k normálnímu rozdělení (viz. další kapitola).
Jestliže náhodná veličina má binomické rozdělení, pak tvar jejího rozložení se blíží k Poissonovu s parametrem λ = n.p, jestliže n je velké a p se blíží k nule. Aproximativně

můžeme tedy binomické rozdělení s velkým n *a malou hodnotou* p *nahradit Poissonovým rozdělením.*

Součet nezávislých proměnných s Poissonovým rozdělením je opět rozdělen podle tohoto rozdělení. Jestliže máme n *pozorování Poissonova rozdělení s parametrem* λ*, pak součet pozorování je možné považovat za pozorování s Poissonovým rozdělením a parametrem* nλ.

Řešené úlohy

Příklad 4.4.1. Předpokládejme, že realitní makléř jedná v průměru s pěti zákazníky za den. Zjistěte jaká je pravděpodobnost, že počet zákazníků za jeden den bude větší než 4.

Řešení: Náhodná veličina X - počet zákazníků přesně splňuje kritéria pro Poissonovo rozdělení. Pravděpodobnostní funkce počtu zákazníků má tedy tvar:

$$p(x)=\frac{5^x}{x!}.e^{-5}$$

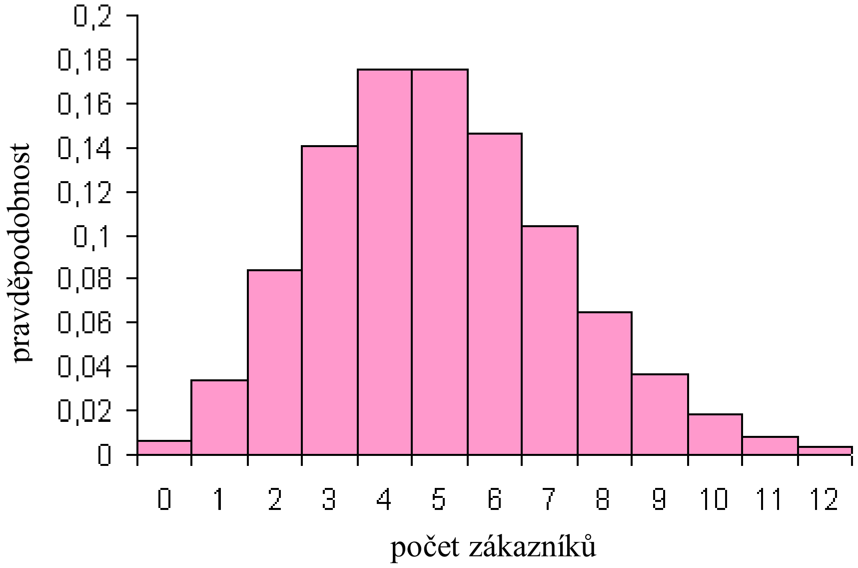
Úlohu nejlépe vyřešíme pomocí opačného jevu:

$$\begin{aligned}P(X>4)&=1-P(X\le 4)=\\&=1-p(0)-p(1)-p(2)-p(3)-p(4)=\\&=1-0,44=0,56\end{aligned}$$

V Excelu bychom výše uvedenou pravděpodobnost vypočetli pomocí funkce POISSON:

$P(X > 4)$ = 1 - POISSON(4;5;1) = 0,56

Poissonovo rozdělení pravděpodobnosti počtu zákazníků:



4.5. Hypergeometrické rozdělení *H*(*N*,*M*,*n*)

Předpokládejme, že náhodný pokus, jehož výsledkům je přiřazena alternativní náhodná veličina $A(p)$, opakujeme n-krát, přičemž jednotlivé pokusy jsou vzájemně závislé (výsledek v libovolném pokusu závisí na předcházejících pokusech) - jedná se tedy o **výběry bez vracení** (opakované pokusy závislé). Pro takto vzniklou náhodnou veličinu X platí:

Definice 4.5.1.

Náhodná veličina X má **hypergeometrické rozdělení** $H(N, M, n)$ právě tehdy, když má pravděpodobnostní funkce tvar:

$$p\ x = \frac{\binom{M}{x}.\binom{N-M}{n-x}}{\binom{N}{n}},$$

kde N je počet prvků základního souboru; M je počet prvků v základním souboru, které mají požadovanou vlastnost; n je počet pokusů a $x = 0, 1, 2, .., n$ je počet vybraných výrobků, které mají zkoumanou vlastnost.

Poznámka

Pravděpodobnostní funkci hypergeometrického rozložení pravděpodobnosti lze snadno odvodit z klasické definice pravděpodobnosti - *viz. kapitola 2.*

Vlastnosti:

- $E(x) = n.\dfrac{M}{N}$
- $D(x) = n.\dfrac{M}{N}.\left(1-\dfrac{M}{N}\right).\left(\dfrac{N-n}{N-1}\right)$

Řešené úlohy

Příklad 4.5.1. Mezi stovkou výrobků je 20 zmetků. Vybereme deset výrobků a sledujeme počet zmetků mezi vybranými.

Řešení: V tomto případě má náhodná veličina X hypergeometrické rozdělení: $X \sim H(100,20,10)$.

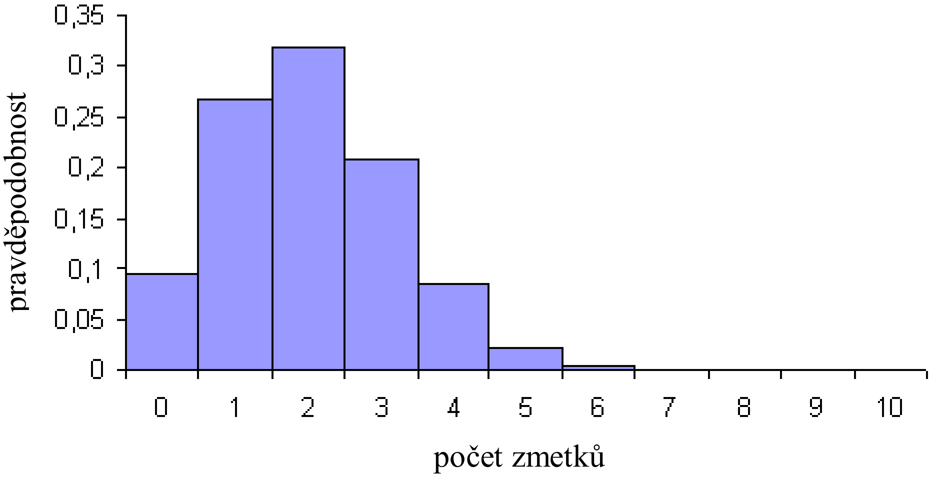
Pravděpodobnostní funkce má tvar:

$$p(x) = \frac{\binom{20}{x}\cdot\binom{80}{10-x}}{\binom{100}{10}}$$

Takže například pravděpodobnost, že mezi deseti vybranými budou 3 zmetky, se vypočte:

$$p(3) = \frac{\binom{20}{3}\cdot\binom{80}{7}}{\binom{100}{10}} \doteq 0{,}209$$

Pravděpodobnostní funkci znázorníme opět graficky:



Úlohy k samostatnému řešení

Diskrétní náhodná veličina

4.1. V zásilce 100 výrobků je 80 výrobků 1. jakosti a 20 výrobků 2. jakosti. Vybíráme třikrát po jednom výrobku a výrobek vždy vracíme zpět. Určete pravděpodobnost, že všechny vybrané výrobky budou 1. jakosti.

4.2. Dlouhodobým pozorováním stavu vody v řece byla určena pravděpodobnost jarní povodně na $\frac{4}{15}$. Určete $E(x)$ a $D(x)$ počtu povodní v nejbližších 100 letech.

4.3. Při výstupní kontrole se z každých 100ks výrobků vybírá 30. Určete střední hodnotu a rozptyl počtu nekvalitních výrobků mezi těmito 30 kusy, je-li zmetkovitost výroby 2 %.

4.4. Za jasných letních nocí můžeme v průměru každých 10 minut vidět "padat hvězdu". Jaká je pravděpodobnost, že během 15 minut uvidíme dvě "padající hvězdy"?

4.5. Trolejbusy odjíždějí ze zastávky v 10 min. intervalech. Cestující může přijít na zastávku v libovolném okamžiku. Určete $E(x)$ a $D(x)$ doby čekání na odjezd trolejbusu.

4.6. Pekárna dodává ráno čerstvé pečivo kdykoliv mezi 5. a 6. hodinou. Jaká je pravděpodobnost, že pečivo bude dodáno mezi 5:30 a 5:45?

4.7. Ke 400 šroubům M10 bylo omylem přimícháno 100 šroubů M8.

a) Jaké bude rozdělení pravděpodobnosti, že při náhodném výběru 5 šroubů bude $m = 1, 2, ..., 5$ šroubů správného rozměru?

b) Pro montáž přístroje potřebuje pracovník 4 šrouby rozměru M10. Jaká je pravděpodobnost, že mezi vybranými 5 šrouby budou alespoň 4 s požadovanými vlastnostmi?

4.8. V dodávce 80 polotovarů je 8 (tj. 10 %) vadných. Náhodně vybereme (najednou, tj. "bez opakování") 5 kusů polotovarů k další kompletaci. Jaká je pravděpodobnost, že mezi vybranými prvky bude maximálně jeden vadný? (řešení v excelu)

4.9. Ke kontrole v továrně je připraveno 100 výrobků. Z nich se náhodně vybírá 20 kusů. Určete střední hodnotu a rozptyl počtu zmetků ve vybraných dvaceti výrobcích, víme-li, že zmetkovitost výroby je 3 %.

4.10. Při výrobě aluminiových odlitků byla zkoumána bublinatost na vymezené ploše odlitků. Zkoumání bylo provedeno na souboru 250 odlitků, u nichž bylo zjištěno celkem 340 bublin. Vyjádřete rozdělení pravděpodobnosti počtu bublin na jednom odlitku.

4.11. Televizor má za 10 000 hodin chodu v průměru 10 poruch. Určete pravděpodobnost poruchy za 200 hodin chodu. Ověřte, zda patřičné binomické rozdělení lze nahradit rozložením Poissonovým.

4.12. Ve skladišti závodu je 5 000 výrobků stejného typu. Pravděpodobnost toho, že daný výrobek nevydrží kontrolní zapojení, je 0,1 %. Najděte pravděpodobnost, že z výrobků na skladě více než dva nevydrží kontrolní zapojení.

4.13. Ve strojírenském závodě se vyrábějí určité součástky, jejichž rozměry mají nahodilé odchylky řídící se normálním zákonem rozložení se směrodatnou odchylkou 4 mm. Výrobky s odchylkou menší než 5 mm se zařazují do vyšší jakostní třídy. Určete střední hodnotu počtu výrobků zařazených do vyšší jakostní třídy z daných 4 výrobků.

4.14. Průměrný počet poruch elektronické aparatury za 10 000 hodin provozu je 10. Určete pravděpodobnost poruchy aparatury za 100 hodin práce.

4.15. Aparatura obsahuje 2 000 stejně spolehlivých součástek, u nichž je pravděpodobnost poruchy $p = 0{,}0005$. Jaká je pravděpodobnost poruchy aparatury, která přestane pracovat i při poruše jediné součástky?

4.16. Pravděpodobnost toho, že výrobek nevydrží zátěž, je 0,001. Najděte pravděpodobnost toho, že z 5 000 výrobků více než jeden nevydrží zatížení. Srovnejte výsledky získané pomocí rozložení binomického a Poissonova.

4.17. Najděte pravděpodobnost toho, že mezi 200 výrobky se vyskytnou více než tři zmetky, když v průměru je zmetkovitost výroby těchto výrobků 1 %.

4.18. Korektura 500 stránek obsahuje 500 nalezených tiskových chyb. Najděte pravděpodobnost toho, že na stránce jsou nejméně tři chyby.

Výsledky úloh k samostatnému řešení

4.1. 0,512

4.2. 26,6; 19,5

4.3. 0,6; 0,416

4.4. 0,251

4.5. 5; 25/3

4.6. 0,25

4.7. $f(x) = C_x(5).0{,}8^x.0{,}2^{5-x}$

4.8. 0,92437, hypergeometrické rozložení

4.9. $p(x) = C_x(3).C_{20-x}(100\text{-}3)$, $n = 20$, $p = 0{,}03$, $x = n.p = 0{,}6$, $\sigma^2 = n.p.q.(N\text{-}n)/(N\text{-}1)=0{,}470$

4.10. $\lambda = 340/250 = 1{,}4$, Poissonovo rozložení

4.11. $p_n = 10 / 10\,000 = 10^{-3}$, $n = 200$, $x = n.p = 0{,}2 \approx n.p.q = 0.1998$, $p(x \neq 0) = 0.181269$

4.12. $x = 5\,000.10^{-3} = 5 = \lambda$, $p(x>2) = 0.875348$

4.13. $3{,}1552 \approx 3$

4.14. $1 - e^{-0,1} = 0{,}095$

4.15. $1 - e^{-1} \approx 0.63$

4.16.
$$1-e^{-5}\sum_{x=0}^{1}\frac{5^x}{x!}=0{,}959572\,,\ 1-\sum\binom{5000}{x}.0{,}001^x.0{,}999^{5000-x}=0{,}959639$$

4.17.
$$1-e^{-2}\sum_{x=0}^{1}\frac{2^x}{x!}=0{,}142876,\ 1-\sum\binom{200}{x}.0{,}01^x.0{,}99^{200-x}=0{,}141965$$

4.18.
$$1-e^{-1}\sum_{x=0}^{2}\frac{1}{x!}=0{,}0803013$$

5. ZÁKLADNÍ TYPY ROZDĚLENÍ PRAVDĚPODOBNOSTI SPOJITÉ NÁHODNÉ VELIČINY

Průvodce studiem

V teto kapitole se seznámíte se základními typy rozložení spojité náhodné veličiny. Vašim úkolem by neměla být pouze základní pasivní znalost a orientace v rozloženích, ale měli byste se také naučit tato rozložení od sebe rozlišovat a bezpečně je rozpoznávat.

Předpokládané znalosti

Pojmy z kombinatoriky, z počtu pravděpodobnosti, derivace, integrál.

Cíle

Cílem této kapitoly je seznámení se základními typy rozložení spojité náhodné veličiny, odvození jejich základních číselných charakteristik.

Výklad

5.1. Rovnoměrné rozdělení *R*(*a*, *b*)

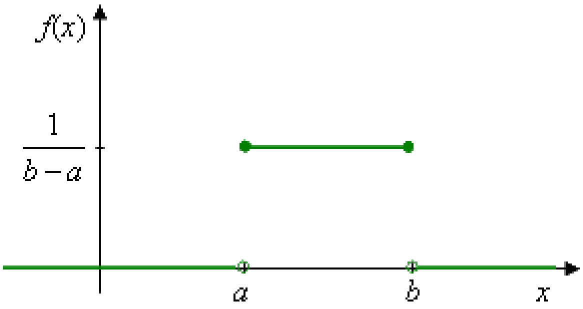
Toto rozdělení má spojitá náhodná veličina X, jejíž realizace vyplňují interval konečné délky a mají stejnou možnost výskytu (např. doba čekání na autobus, na výrobek u automatické linky, ...).

Definice 5.1.1.

Náhodná veličina X má **rovnoměrné rozdělení** $R(a,b)$ právě tehdy, když má hustota pravděpodobnosti rovnici:

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

Graf hustoty pravděpodobnosti:



Distribuční funkce je ve tvaru:

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

Poznámka

Vyjádření distribuční funkce lze snadno odvodit ze základní vlastnosti distribuční funkce a hustoty pravděpodobnosti:

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

Tudíž:

$x\in(-\infty,a)$:

$$F(x)=\int_{-\infty}^{x} 0\,dt=0$$

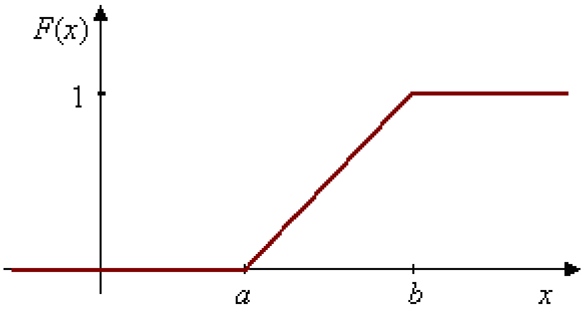
$x\in\langle a,b\rangle$:

$$F(x)=\int_{a}^{x}\frac{1}{b-a}\,dt=\frac{1}{b-a}.\left[t\right]_a^x=\frac{x-a}{b-a}$$

$x\in(b,\infty)$:

$$F(x)=\int_{a}^{b}\frac{1}{b-a}\,dt+\int_{b}^{x}0\,dt=\frac{b-a}{b-a}=1$$

Graf distribuční funkce:



Vlastnosti:

- $E(x) = \dfrac{a+b}{2}$
- $D(x) = \dfrac{(b-a)^2}{12}$

Tyto vlastnosti můžeme opět velmi jednoduše odvodit:

$$E(x) = \mu = \int_a^b x.f(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2.(b-a)} = \frac{a+b}{2}$$

$$D(x) = \mu_2 - \mu^2 = \int_a^b x^2.f(x)\,dx - \mu^2 = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b - \mu^2 =$$

$$= \frac{b^3-a^3}{3.(b-a)} - \left(\frac{a+b}{2}\right)^2 = \ldots = \frac{(b-a)^2}{12}$$

Řešené úlohy

Příklad 5.1.1. Tramvajová linka číslo 8 odjíždí v dopoledních hodinách ze zastávky každých 10 minut. Vypočtěte pravděpodobnost, že na ni budete dopoledne čekat déle než 7 minut.

Řešení: Doba čekání je náhodná veličina *X*, která má rovnoměrné rozdělení pravděpodobnosti - v našem případě *R*(0,10). Distribuční funkce má tedy tvar:

$$F(x) = \begin{cases} 0 & \text{pro } x \in (-\infty, 0) \\ \dfrac{x}{10} & \text{pro } x \in \langle 0, 10 \rangle \\ 1 & \text{pro } x \in (10, \infty) \end{cases}$$

Hledaná pravděpodobnost:

$$P(X>7)=P(7<X<\infty)=F(\infty)-F(7)=1-\frac{7}{10}=\frac{3}{10}$$

5.2. Exponenciální rozdělení *E*(λ)

Toto rozdělení má spojitá náhodná veličina *X*, která představuje dobu čekání do nastoupení (poissonovského) náhodného jevu, nebo délku intervalu (časového nebo délkového) mezi takovými dvěma jevy (např. doba čekání na obsluhu, vzdálenost mezi dvěma poškozenými místy na silnici).

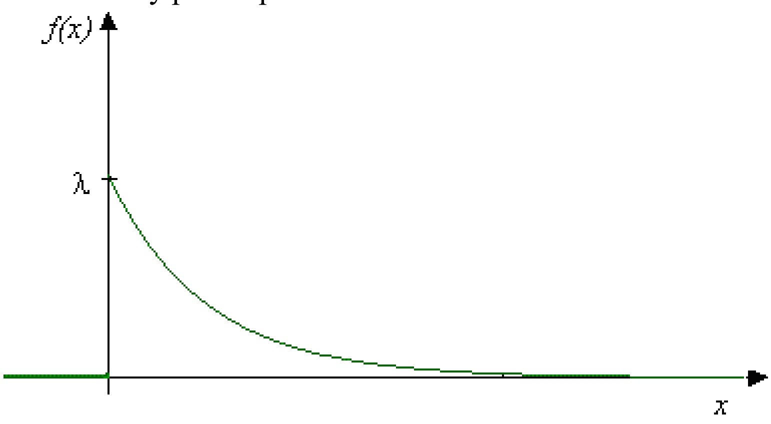
Závisí na parametru λ, což je převrácená hodnota střední hodnoty doby čekání do nastoupení sledovaného jevu.

Definice 5.2.1.

Náhodná veličina *X* má **exponenciální rozdělení** *E*(λ) právě tehdy, když je hustota pravděpodobnosti dána vztahem:

$$f(x)=\begin{cases}0 & \text{pro } x<0\\ \lambda.e^{-\lambda x} & \text{pro } x\geq 0\end{cases}$$

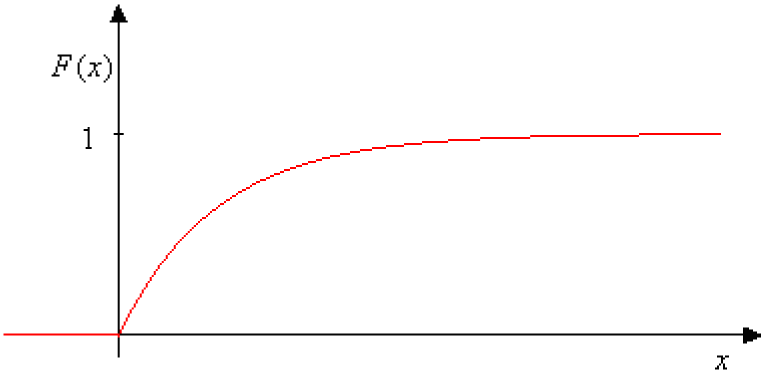
Graf hustoty pravděpodobnosti:



Distribuční funkce:

$$F(x)=\begin{cases}0 & \text{pro } x<0\\ 1-e^{-\lambda x} & \text{pro } x\geq 0\end{cases}$$

Graf distribuční funkce:



Vlastnosti:

- $E(x) = \dfrac{1}{\lambda}$
- $D(x) = \dfrac{1}{\lambda^2}$

Poznámka

Tvar distribuční funkce, stejně jako vlastnosti exponenciálního rozdělení, lze odvodit obdobně jednoduchým způsobem, jako u rovnoměrného rozdělení.

Řešené úlohy

Příklad 5.2.1. Doba čekání hosta na pivo je v restauraci U Lva průměrně 5 minut. Určete:

a) hustotu pravděpodobnosti náhodné veličiny, která je dána dobou čekání na pivo

b) pravděpodobnost, že budeme čekat na pivo déle než 12 minut

c) dobu čekání, během které bude zákazník obsloužen s pravděpodobností 0,9

Řešení: Jedná se tedy o exponenciální rozložení pravděpodobnosti:

a) Hustota pravděpodobnosti:

$$f(x) = \begin{cases} 0 & \text{pro } x < 0 \\ \frac{1}{5}.e^{\frac{-1}{5}x} & \text{pro } x \geq 0 \end{cases}$$

b) Distribuční funkce:

$$F(x) = \begin{cases} 0 & \text{pro } x < 0 \\ 1 - e^{\frac{-1}{5}x} & \text{pro } x \geq 0 \end{cases}$$

Hledaná pravděpodobnost:

$$P(X>12)=P(12<X<\infty)=F(\infty)-F(12)=$$

$$=1-\left(1-e^{\frac{-1}{5}.12}\right)=e^{-\frac{12}{5}}\doteq 0,0907$$

c) Hledanou dobu čekání označíme *t*. Platí:

$$P(0<X\le t)=0,9$$

$$F(t)-F(0)=0,9$$

$$1-e^{\frac{-1}{5}.t}-0=0,9$$

$$e^{\frac{-1}{5}.t}=0,1$$

$$-\tfrac{1}{5}t=\ln 0,1$$

$$t=-5.\ln 0,1$$

$$t\doteq 11,51\,\text{minut}$$

$$\text{t}\doteq 11\,\text{minut 30 sekund}$$

5.3. Normální rozdělení *N*(μ, σ²)

Označováno též *obecné normální rozdělení* či *Gaussovo rozdělení* (v anglicky psané literatuře nazývané *rozdělení zvonovitého tvaru* - bell curve).

Je velmi důležité, neboť:

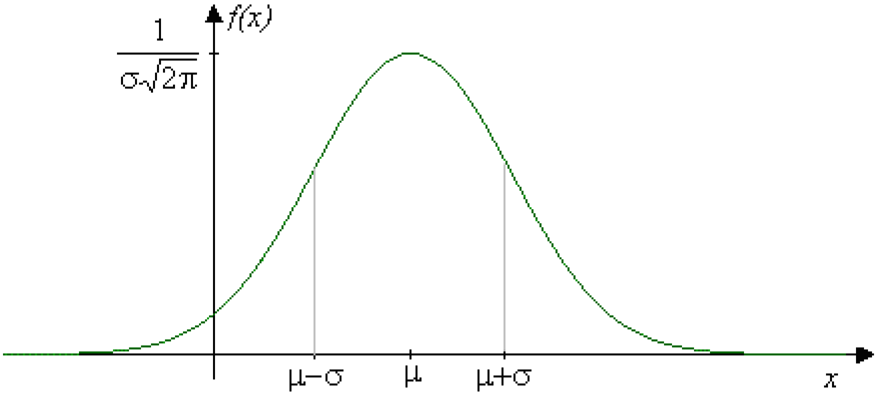
- nejčastěji se vyskytuje
- mnoho jiných rozdělení se mu blíží
- řada jiných rozdělení se jím dá nahradit

Definice 5.3.1.

Náhodná veličina *X* má **normální rozdělení** *N*(μ, σ²) právě tehdy, když má hustota pravděpodobnosti tvar:

$$f(x)=\frac{1}{\sigma.\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\quad \text{pro } x\in(-\infty,\infty)$$

Grafem hustoty pravděpodobnosti je tzv. Gaussova (Gaussova-Laplaceova) křivka:

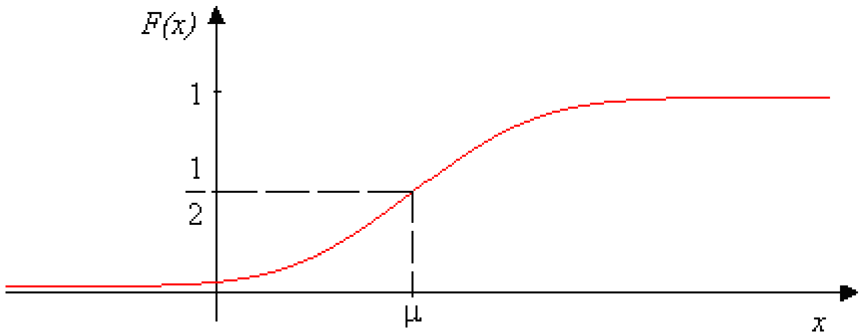


Z obrázku je patrné, že parametr μ (střední hodnota) určuje, kde má křivka maximum. Parametr σ (směrodatná odchylka) naproti tomu určuje, jak jsou po obou stranách od hodnoty μ vzdáleny inflexní body, tedy jak je křivka roztažena do šířky.

Distribuční funkce:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma.\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \quad \text{pro } x \in (-\infty, \infty)$$

Graf distribuční funkce:



Poznámka

Pomocí křivky normálního rozdělení popsal v roce 1773 matematik Abraham de Moivre limitní chování binomického rozdělení, když se snažil aproximovat výpočty jednotlivých pravděpodobností binomického rozdělení pro velká n. *Rozdělení, které Moivre pro tento účel navrhl, se nakonec ukázalo být důležitější než výchozí binomické rozdělení. V roce 1812 odvodil nezávisle* na *Moivreovi normální rozdělení francouzský matematik Pierre Laplace. Jak Laplace, tak Karl Friedrich Gauss prezentovali toto rozdělení jako zákon chyb a* *používali ho pro interpretaci astronomických a geodetických měření, výsledků hazardních her a přesnosti dělostřelecké střelby.*

Řešené úlohy

Příklad 5.3.1. Jaká je pravděpodobnost, že náhodná veličina X, která má rozdělení N(10, 9), nabude hodnoty

a) menší než 16,

b) větší než 10,

c) v mezích od 7 do 22?

Řešení:

a)

$$P(X<16)=P(-\infty<X<16)=F(16)-F(-\infty)=F(16)$$

Zjistit, čemu je rovna distribuční funkce pro hodnotu 16 můžeme několika způsoby. V příští kapitole si ukážeme, že náhodnou veličinu můžeme převést na normované normální rozdělení $N(0, 1)$, jehož hodnoty jsou v tabulkách. Máme-li ale k dispozici např. program Excel, můžeme hodnotu vypočíst pomocí předdefinované funkce NORMDIST:

$P(X < 16) = F\{16) =$ NORMDIST(16;10;3;1) = 0,97725

První parametr v závorce je hodnota, jejíž distribuční funkci počítáme, druhý je střední hodnota daného normálního rozdělení, třetí parametr je směrodatná odchylka daného rozdělení a poslední parametr je pravdivostní hodnota 1, kterou zadáme vždy, když chceme vypočítat hodnotu distribuční funkce.

b) $P(X > 10) = P(10 < X < \infty) = 1 - F(10) =1 -$ NORMDIST(10;10;3;1) = 0,5

c) $P(7 < X < 22) =$ NORMDIST(22;10;3;1) - NORMDIST(7;10;3;1) = 0,8413

5.4. Normované normální rozdělení *N*(0, 1)

Jedná se o speciální případ obecného normálního rozložení, kdy $\mu = 0$, $\sigma^2 = 1$.

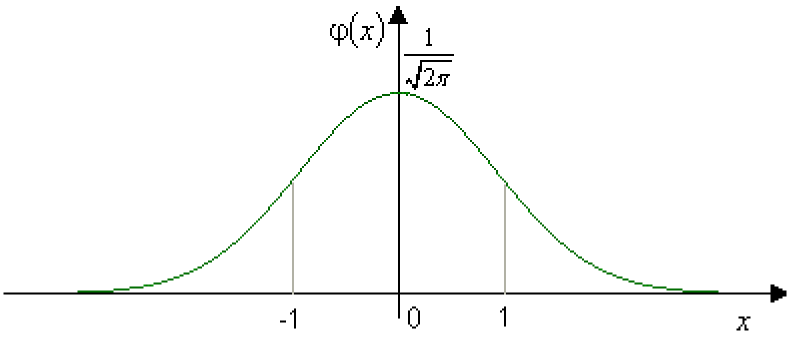
V tomto případě označujeme hustotu pravděpodobnosti:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}x^2} \quad \text{pro } x\in(-\infty,\infty)$$

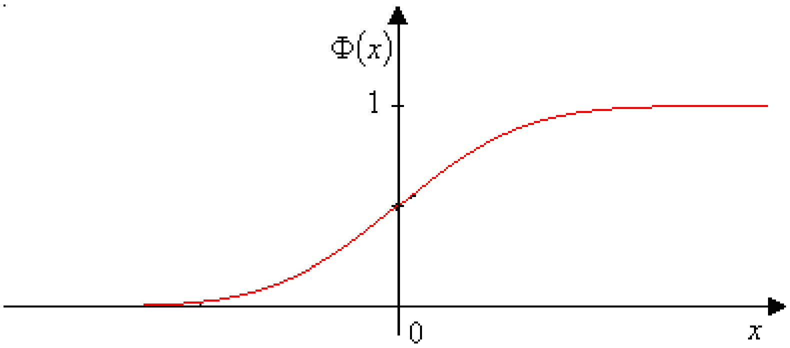
Distribuční funkci u tohoto rozdělení označujeme:

$$\Phi\ x\ = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt \quad \text{pro } x \in\ -\infty, \infty$$

Graf hustoty pravděpodobnosti:



Graf distribuční funkce:



Užitečnost normovaného normálního rozdělení spočívá v tom, že vybrané hodnoty distribuční funkce tohoto rozdělení najdeme v tabulkách, které bývají součástí každé učebnice statistiky. Vztah mezi normovaným normálním rozdělením $N(0,1)$ a obecným normálním rozdělením $N(\mu, \sigma^2)$ vyjadřuje následující věta:

Věta 5.4.1.

Má-li spojitá náhodná veličina X obecné normální rozdělení $N(\mu, \sigma^2)$ s hustotou pravděpodobnosti: $f(x) = \frac{1}{\sigma.\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ pro $x \in (-\infty, \infty)$,

pak náhodná veličina $T = \frac{X-\mu}{\sigma}$ má normované normální rozdělení $N(0,1)$ s hustotou pravděpodobnosti:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}t^2} \quad \text{pro } t \in (-\infty, \infty)$$

Důkaz: Zavedeme-li do vztahu:

$$P(X < x_0) = \frac{1}{\sigma.\sqrt{2\pi}}.\int_{-\infty}^{x_0} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \text{ substituci:}$$

$$T = \frac{X-\mu}{\sigma}, dt = \frac{dx}{\sigma}, \text{ dostáváme:}$$

$$P(T < t_0) = \frac{1}{\sqrt{2\pi}}.\int_{-\infty}^{t_0} e^{-\frac{1}{2}t^2} dt, \text{ kde } t_0 = \frac{x_0-\mu}{\sigma}.$$

Poznámka

V tabulkách nalezneme pouze hodnoty distribuční funkce pro nezáporné t. *Chceme-li určit distribuční funkci pro* t < 0*, využijeme vlastností distribuční funkce normovaného normálního rozdělení a můžeme lehce odvodit, že* $\Phi(-t) = 1 - \Phi(t)$

Řešené úlohy

Příklad 5.4.1. Použijeme zadání příkladu 5.3.1., přičemž tento příklad vyřešíme převedením daného normálního rozdělení $N(10, 9)$ na normované normální rozdělení $N(0, 1)$ substitucí z předchozí věty 5.4.1.

Řešení:

a)

$$P(X<16)=P(-\infty<X<16)=F(16)-F(-\infty)=$$

$$=F(16)=\Phi\left(\frac{16-10}{3}\right)=\Phi(2)=0{,}97725$$

b) $P(X > 10) = P(10 < X < \infty) = 1 - F(10) = 1 - \Phi(0) = 0{,}5$

c) $P(7 < X < 22) = \Phi(4) - \Phi(-1) = = \Phi(4) - 1 + \Phi(1) = 0{,}8413$

Všechny hodnoty jsou dosazené z tabulky distribuční funkce normálního rozdělení.

Příklad 5.4.2. Určete pravděpodobnost, že náhodná veličina X s normálním rozdělením $N(\mu, \sigma^2)$ nabude hodnot z intervalu

a) $(\mu-\sigma,\mu+\sigma)$

b) $(\mu-2\sigma,\mu+2\sigma)$

c) $(\mu-3\sigma,\mu+3\sigma)$

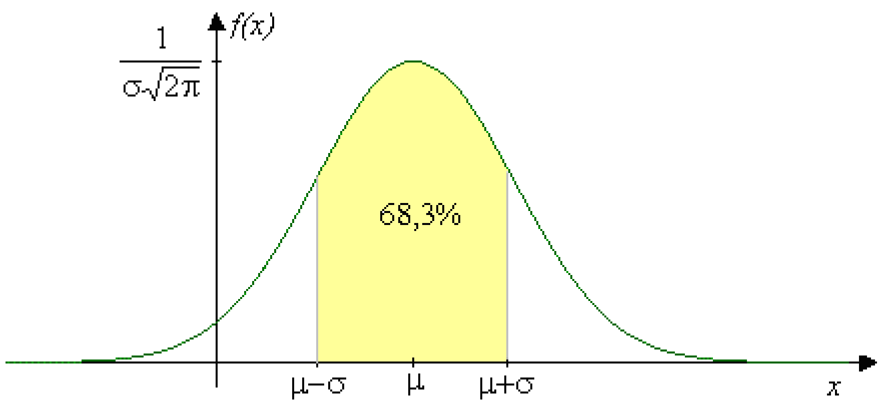
Řešení:

a)

$$P(\mu-\sigma<X<\mu+\sigma)=F(\mu+\sigma)-F(\mu-\sigma)=\Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right)-\Phi\left(\frac{\mu-\sigma-\mu}{\sigma}\right)=$$

$$=\Phi(1)-\Phi(-1)=\Phi(1)-(1-\Phi(1))=2.\Phi(1)-1\approx 0{,}683$$

Grafické znázornění:



b)

$$P(\mu-2\sigma<X<\mu+2\sigma)=F(\mu+2\sigma)-F(\mu-2\sigma)=$$

$$=\ldots=2.\Phi(2)-1\approx 0{,}955$$

c)

$$P(\mu-3\sigma < X < \mu+3\sigma) = F(\mu+3\sigma) - F(\mu-3\sigma) =$$
$$= \ldots = 2.\Phi(3) - 1 \doteq 0{,}997$$

Poznámka

*Výsledek příkladu 5.4.2*c. *je znám pod názvem* **pravidlo 3σ**. *Vyjadřuje skutečnost, že náhodná veličina s obecným normálním rozdělením* N(μ, σ^2) *nabude hodnot z intervalu* ($\mu-3\sigma,\mu+3\sigma$) *s pravděpodobností 97,7 %.*

5.4.1. Aproximace binomického rozdělení

U binomického rozdělení může být pro velká n obtížný výpočet kombinačních čísel.

Jak už bylo řečeno, binomické rozdělení lze aproximovat Poissonovým a to v případě, že $p < 0{,}3$ nebo $p > 0{,}7$:

Bi(*n*, *p*) $\longrightarrow$ ***Po*(λ)**, kde **λ = *n.p***

Jestliže $p \in \langle 0{,}3; 0{,}7 \rangle$:

Bi(*n*, *p*) $\longrightarrow$ ***N*(μ, σ²)**, kde **μ = *n.p*, σ² = *n.p*(1 - *p*)**

Řešené úlohy

Příklad 5.4.3 Házíme 100 krát mincí. Jaká je pravděpodobnost, že lev padne aspoň 50 krát?

Řešení: *X*...počet padnutí lva

Náhodná veličina *X* má binomické rozdělení, neboť házení mincí jsou opakované pokusy - nezávislé. Problém při řešení tohoto příkladu může nastat ve chvíli, kdy nemáme k dispozici žádný software, který by dokázal počítat hodnoty binomického rozdělení - museli bychom tedy ručně sčítat 51 hodnot pravděpodobnostní funkce binomického rozdělení mezi 50 a 100.

Máme-li k dispozici alespoň statistické tabulky, můžeme řešit pomocí normálního rozdělení: $N(\mu, \sigma^2)$, kde:

$\mu = n.p = 50$

$\sigma^2 = n.p.(1 - p) = 25$

Takže:

$P(X = 50 \vee 51 \vee 52 \vee \ldots \vee 100) = 1 - P(X < 50) = 1 - F(50) = 1 - \Phi(0) = 0{,}5$

5.5. Některá další rozdělení

5.5.1. Weibullovo rozdělení $W(\delta, c)$

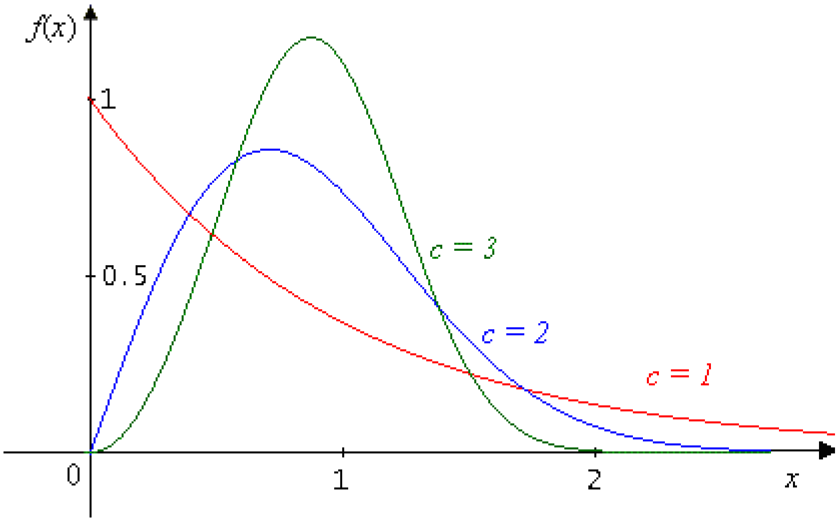
Toto rozdělení má spojitá náhodná veličina, která představuje dobu života (bezporuchovosti) technických zařízení, kterým nevyhovuje exponenciální. To jest tam, kde se projevuje mechanické opotřebení nebo únava materiálu.

Parametr δ závisí na materiálu, namáhání a podmínkách užívání ($\delta > 0$); $c > 0$.

Funkce hustoty pravděpodobnosti:

$$f(x) = \begin{cases} 0 & \text{pro } x \le 0 \\ \dfrac{c.x^{c-1}}{\delta^c}.e^{-\left(\frac{x}{\delta}\right)^c} & \text{pro } x > 0 \end{cases}$$ (pro $c = 1$ dostaneme exponenciální rozdělení $E(\delta)$)

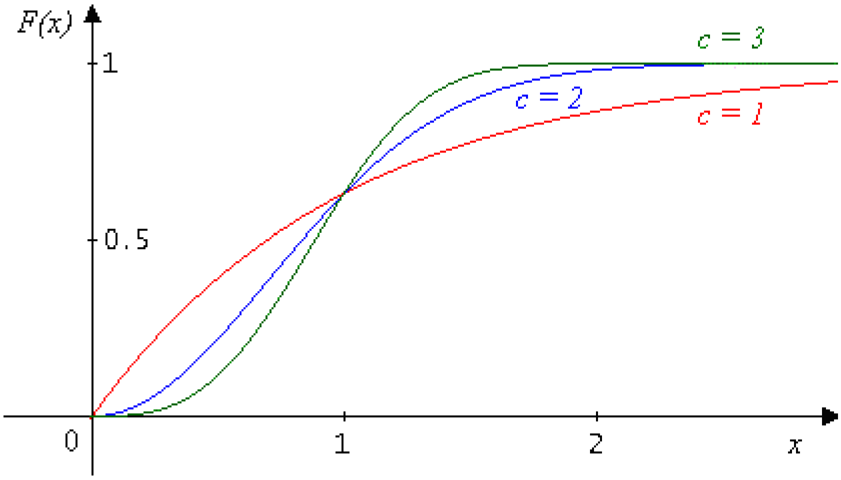
Grafické znázornění hustoty pravděpodobnosti pro $\delta = 1$ a různé hodnoty c:



Distribuční funkce:

$$F(x) = \begin{cases} 0 & \text{pro } x \le 0 \\ 1 - e^{-\left(\frac{x}{\delta}\right)^c} & \text{pro } x > 0 \end{cases}$$

Grafické znázornění distribuční funkce pro δ = 1 a různé hodnoty c:



5.5.2. Pearsonovo rozdělení χ_n^2

χ_n^2 ... čteme chí kvadrát s n stupni volnosti

Užití: Jestliže n nezávislých veličin $X_1,\dots,X_n$ má rozdělení $N(0, 1)$, pak veličina $X=X_1^2+X_2^2+\dots+X_n^2$ má Pearsonovo rozdělení.

Hustota pravděpodobnosti:

$$f(x)=\begin{cases}\dfrac{x^{\frac{n}{2}-1}.e^{-\frac{x}{2}}}{2^{\frac{n}{2}}.\Gamma\left(\frac{n}{2}\right)} & \text{pro } x>0\\ 0 & \text{pro } x\le 0\end{cases}$$

$\Gamma(x)$...gama funkce definovaná pro $x > 1$ vztahem: $\Gamma(x)=\int_0^\infty e^{-t}.t^{x-1}dt$

5.5.3. Studentovo rozdělení t_n

Užití: Jsou-li X_1,X_2 dvě nezávislé náhodné proměnné, kde X_1 se řídí rozložením $N(0, 1)$ a X_2 rozložením χ_n^2, pak náhodná veličina $T=\dfrac{X_1}{\sqrt{X_2}}.\sqrt{n}$ má Studentovo rozložení s n stupni volnosti.

$$f(x)=\frac{1}{\sqrt{n\pi}}.\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Úlohy k samostatnému řešení

Spojitá náhodná veličina

5.1. Náhodná veličina má hustotu pravděpodobnosti:

$$f(x) = \begin{cases} 0{,}1.e^{-0{,}1x} & \text{pro } x > 0 \\ 0 & \text{pro } x \le 0 \end{cases}.$$

Určete její střední hodnotu a rozptyl.

5.2. Náhodná veličina X má rozdělení $N(0,\ 1)$. Určete:

a) $P(X < 2{,}31)$

b) $P(X < -1{,}1)$

c) $P(-0{,}41 < X < 2{,}92)$

5.3. Náhodná veličina X má rozdělení $N(2,\ 9)$. Určete:

a) $P(X < 5)$

b) $P(X < -1)$

c) $P(0 < X < 2{,}33)$

5.4. Náhodná veličina má rozdělení pravděpodobnosti: *a*) $N(0,\ 1)$

b) $N(0{,}4)$

c) $N(1{,}4)$

Určete v případě *a*) $P(|X| < 0{,}7)$; b), c) $P(X < -0{,}5)$. Sestrojte graf $f(x)$, $F(x)$ a vypočtené pravděpodobnosti znázorněte.

5.5. Jaká je pravděpodobnost, že náhodná veličina X, která má rozdělení N(10; 9), nabude hodnoty

a) menší než 16,

b) větší než 10,

c) v mezích od 7 do 22?

5.6. Jaká je pravděpodobnost, že při 100 hodech mincí padne lev aspoň čtyřicetkrát a maximálně padesátkrát?

5.7. Jaká je pravděpodobnost, že při 60 hodech kostkou nepadne 6 ani jednou?

5.8. Basketbalista dá koš s pravděpodobností 0,6. Jaká je pravděpodobnost, že při 60 hodech bude úspěšný aspoň třicetkrát a nejvýše čtyřicetkrát?

5.9. Měření je zatíženo chybou -0,3 cm. Náhodné chyby měření mají normální rozdělení pravděpodobnosti se směrodatnou odchylkou $\sigma = 0,5$ cm. Jaká je pravděpodobnost, že chyba měření nepřekročí v absolutní hodnotě trojnásobek směrodatné odchylky?

5.10. Váha v uhelných skladech váží s chybou 30 kg, přičemž snižuje váhu. Náhodné chyby mají normální rozdělení pravděpodobnosti se $\sigma = 100$ kg. Jaká je pravděpodobnost, že chyba zjištěné váhy nepřekročí v absolutní hodnotě 90 kg?

5.11. Kolik procent hodnot náhodné veličiny X s rozdělením $N(0, 1)$ leží mimo interval (-2, 2)?

5.12. Jakou je nutno stanovit toleranci, aby pravděpodobnost, že průměr pískového zrna překročí toleranční hranici, byla maximálně 0,45326, jestliže odchylky od středu tolerance (v 10^{-2} mm) mají normální rozdělení $N(0, 144)$.

Výsledky úloh k samostatnému řešení

5.1. 10; 100

5.2. 0,98956; 0,13567; 0,65735

5.3. 0,84134; 0,15866; 0,29130

5.4. 0,51608; 0,40129; 0,22663

5.5. *a*) 0,97725, *b*) 0,5, *c*) 0,84131

5.6. 0,47725

5.7. 1,77.10-5 - pomocí binomického rozdělení;
4,34.10-5 pomocí Poissonova rozdělení

5.8. 0,84

5.9. 0,99164

5.10. 0,61068

5.11. 4,55

5.12. $7{,}2.10^{-2}$

6. NÁHODNÝ VEKTOR

Průvodce studiem

V počtu pravděpodobnosti i v matematické statistice se setkáváme nejen s náhodnými veličinami, jejichž hodnotami jsou reálná čísla, ale i s takovými, jejichž hodnotami jsou uspořádané *n*-tice reálných čísel - např. měříme-li u výrobků několik kvantitativních charakteristik. V těchto případech musíme zavést pojem náhodného vektoru.

Předpokládané znalosti

Pojmy z kombinatoriky, pravděpodobnosti, znalosti z kapitoly náhodná veličina, znalost parciálních derivací, dvojného integrálu.

Cíle

Cílem této kapitoly je objasnit pojmy náhodný vektor, pravděpodobnostní funkce, hustota pravděpodobnosti, distribuční funkce, marginální funkce náhodného vektoru, charakteristiky náhodného vektoru - kovariance, koeficient korelace.

Výklad

6.1. Náhodný vektor - popis

Definice 6.1.1.

Uspořádaná n-tice náhodných veličin $X_1,X_2,...,X_n$ se nazývá ***n*-rozměrný náhodný vektor** (n-rozměrná náhodná veličina) a značí se: $X = (X_1,X_2,...,X_n)$.

$X_1,X_2,...,X_n$ - složky náhodného vektoru

Poznámky

Pro zjednodušení budeme hovořit o dvourozměrném náhodném vektoru $X=(X_1, X_2)$ nebo (X, Y).

Budeme se zabývat pouze náhodnými vektory, jejichž všechny složky jsou buď diskrétní náhodné veličiny nebo spojité náhodné veličiny.

Rozdělení pravděpodobnosti náhodného vektoru popisujeme stejně jako u náhodné veličiny pomocí frekvenční funkce (u diskrétní náhodné veličiny - pravděpodobnostní funkce, u spojité náhodné veličiny - hustota pravděpodobnosti) nebo distribuční funkce:

6.1.1. Distribuční funkce náhodného vektoru (X, Y)

Definice 6.1.2.

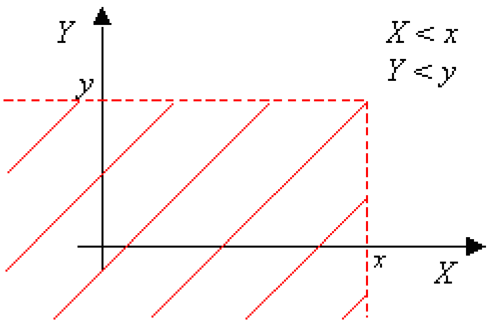
Sdružená (simultánní) **distribuční funkce náhodného vektoru (X, Y)** je reálná funkce $F(x, y)$ definovaná vztahem:

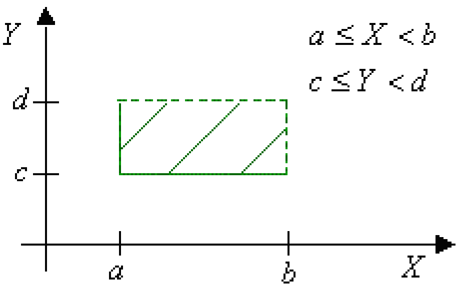
$F(x, y) = P(X < x, Y < y)$

Vlastnosti distribuční funkce:

1. $0 \le F(x,y) \le 1$
2. $F(-\infty,y) = F(x,-\infty) = F(-\infty,-\infty) = 0$; $F(\infty,\infty) = 1$
3. $F(x,y)$ je neklesající funkce
4. $F(x,y)$ je funkce spojitá zleva
5. $P(a \le X < b; c \le Y < d) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$

Grafické vyjádření:





6.1.2. Frekvenční funkce náhodného vektoru (*X*, *Y*)

Diskrétní náhodný vektor

Definice 6.1.3.
Sdružená (simultánní) **pravděpodobnostní funkce náhodného vektoru (*X*, *Y*)** je funkce dána vztahem:
$p(x, y) = P(X = x, Y = y)$

Vlastnosti pravděpodobnostní funkce:

1. $0 \leq p(x_i, y_j) \leq 1$
2. $\sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i, y_j) = 1$
3. $F(x, y) = \sum_{x_i < x}\sum_{y_j < y} p(x_i, y_j)$

Poznámka
Všechny tři vlastnosti jsou obdobné vlastnostem pravděpodobnostní funkce jednorozměrné náhodné veličiny.

Užití:

obecně tabulka				
$X\backslash Y$	y_1	y_2	y_3	$P(X=x_i)$
x_1	$p(x_1,y_1)$	$p(x_1,y_2)$	$p(x_1,y_3)$	$P(X=x_1) = p(x_1,y)$
x_2	$p(x_2,y_1)$	$p(x_2,y_2)$	$p(x_2,y_3)$	$P(X=x_2) = p(x_2,y)$
$P(Y=y_i)$	$P(Y=y_1) = p(x,y_1)$	$P(Y=y_2) = p(x,y_2)$	$P(Y=y_3) = p(x,y_3)$	1

konkrétní příklad tabulky

$X\backslash Y$	**0**	**1**	**2**	$P(X=x_i)$
0	0,42	0,12	0,06	0,6
1	0,28	0,08	0,04	0,4
$P(Y=y_i)$	0,7	0,2	0,1	1

Spojitý náhodný vektor

Definice 6.1.4.

Sdružená (simultánní) **hustota pravděpodobnosti náhodného vektoru (*X, Y*)** je funkce daná vztahem:

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Vlastnosti hustoty pravděpodobnosti:

1. $F(x,y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,y)\,dxdy$

2. $P(a\le X<b, c\le Y<d)=\int\limits_{a}^{b}\int\limits_{c}^{d} f(x,y)\,dxdy$

3. $f(x,y)\ge 0$

4. $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)\,dxdy=1$

Řešené úlohy

Příklad 6.1.1. Najděte konstantu c tak, aby funkce:

$$f(x,y)=\begin{cases} c\dfrac{x^2}{1+y^2} & \text{pro } 2\le x\le 3, 0\le y\le 1 \\ 0 & \text{jinde} \end{cases}$$

byla hustotou pravděpodobnosti nějakého náhodného vektoru (X,Y)

Řešení:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c\frac{x^2}{1+y^2}\,dxdy = 1$$

$$c.\int_2^3 dx\int_0^1 \frac{x^2}{1+y^2}\,dy = 1$$

$$c.\int_2^3 dx\left[x^2.\operatorname{arctg} y\right]_0^1 = 1$$

$$c.\int_2^3 \frac{\pi}{4}x^2 dx = 1$$

$$\frac{\pi}{4}c.\left[\frac{x^3}{3}\right]_2^3 = 1$$

$$\frac{\pi}{4}c.\left(9-\frac{8}{3}\right) = 1$$

$$c = \frac{12}{19\pi}$$

Kromě rozdělení vektoru (*X*, *Y*) nás budou i nadále zajímat rozdělení jednotlivých náhodných veličin *X* a *Y*, kterým budeme říkat **marginální rozdělení,** a rozdělení těchto veličin za jistých podmínek - **podmíněná rozdělení**:

6.1.3. Marginální rozdělení pravděpodobnosti

Definice 6.1.5.

Marginální (okrajové) pravděpodobnostní funkce náhodné veličiny *X* nebo *Y* jsou dány vztahy:

$p_1(x) = P(X = x) = \sum_y p \;\; x, y$

$p_2(y) = P(Y = y) = \sum_x p \;\; x, y$

Marginální (okrajové) hustoty pravděpodobnosti náhodné veličiny *X* nebo *Y* jsou dány vztahy:

$$f_1 \;\; x \;\; = \int_{-\infty}^{\infty} f \;\; x, y \;\; dy$$

$$f_2 \;\; y \;\; = \int_{-\infty}^{\infty} f \;\; x, y \;\; dx$$

Marginální (okrajové) distribuční funkce náhodné veličiny X nebo Y jsou dány vztahy:

$F_1(x) = P(X < x) = F(x, \infty)$

$F_2(y) = P(Y < y) = F(\infty, y)$

6.1.4. Podmíněné rozdělení pravděpodobnosti

Definice 6.1.6.

Podmíněná pravděpodobnostní funkce $p(x/y)$ náhodné veličiny X za podmínky, že náhodná veličina Y nabyla hodnoty y, je:

$$p(x/y) = \frac{p(x,y)}{p_2(y)};\quad p_2(y) \neq 0$$

Podmíněná hustota pravděpodobnosti:

$$f(x/y) = \frac{f(x,y)}{f_2(y)};\quad f_2(y) \neq 0$$

Podmíněná distribuční funkce:

$$F(x/y) = \frac{\sum\limits_{x<x_i} p(x_i, y)}{p_2(y)} \text{ ... pro diskrétní náhodný vektor } \quad p_2(y) \neq 0$$

$$F(x/y) = \frac{1}{p_2(y)} \int_{-\infty}^{x} f(t,y)\,dt \text{ ... pro spojitý náhodný vektor } \quad p_2(y) \neq 0$$

Řešené úlohy

Příklad 6.1.2. Studenti z jedné studijní skupiny byli na zkoušce z matematiky a fyziky s těmito výsledky (první hodnota v uspořádané dvojici označuje výsledek studenta z matematiky, druhá z fyziky):

(1,1), (1,2), (1,3), (2,2), (2,3), (2,3), (3,2), (3,2), (3,3), (3,3), (3,3), (3,3), (3,3), (3,4), (3,4), (4,3), (4,3), (4,4), (4,4), (4,4).

1. Vytvořte pravděpodobnostní tabulku náhodného vektoru, jehož složka X bude znamenat výsledky u zkoušky z matematiky a složka Y bude znamenat výsledky u zkoušky z fyziky

2. Určete jeho marginální pravděpodobnostní funkce $p_1(x)$, $p_2(y)$

3. Určete jeho distribuční funkci $F(x,y)$

4. Zjistěte jeho podmíněné pravděpodobnosti $p(x/y)$

Řešení:

ad 1.

$X\backslash Y$	1	2	3	4
1	0,05	0,05	0,05	0
2	0	0,05	0,1	0
3	0	0,1	0,25	0,1
4	0	0	0,1	0,15

ad 2.

Hodnoty v prvním řádku a prvním sloupci jsou hodnoty, kterých mohou nabývat náhodné veličiny X, Y. Ostatní čísla v tabulce znamenají pravděpodobnosti všech možných dvojic, např. $p\;1,1\; = \frac{1}{20} = 0,05$ (hodnota v druhém řádku a druhém sloupci tabulky) vznikla jako jediná možnost (1, 1) ze všech dvaceti možností.

$X\backslash Y$	1	2	3	4	$p_1(x_i)$
1	0,05	0,05	0,05	0	0,15
2	0	0,05	0,1	0	0,15
3	0	0,1	0,25	0,1	0,45
4	0	0	0,1	0,15	0,25
$p_2(y_j)$	0,05	0,2	0,5	0,25	**1**

Hodnoty marginální pravděpodobnostní funkce $p_1(x_i)$ jsou vždy součty všech pravděpodobností v daném řádku, např.:

$p_1(3) = 0 + 0,1 + 0,25 + 0,1 = 0,45$. Obdobně nalezneme ve sloupcích hodnoty $p_2(y_j)$.

Zvýrazněné číslo musí být vždy rovno jedné, je to součet všech hodnot $p_1(x_i)$ nebo $p_2(y_j)$, tedy vlastně součet všech pravděpodobností náhodného vektoru.

ad 3.

$F(x,y)$

$X\backslash Y$	1	2	3	4	5
1	0	0	0	0	0
2	0	0,05	0,1	0,15	0,15
3	0	0,05	0,15	0,3	0,3
4	0	0,05	0,25	0,65	0,75
5	0	0,05	0,25	0,75	1

postup při výpočtu, např.:

$F(3,3) = P(X<3,Y<3) = p(1,1) + p(1,2) + p(2,1) + p(2,2) = 0,15$

Všimněte si, že hodnoty v posledním sloupci odpovídají hodnotám marginální distribuční funkce $F_1(x)$ a hodnoty v posledním řádku hodnotám $F_2(y)$

ad 4.

$p(x/y)$

$X\backslash Y$	1	2	3	4
1	1	0,25	0,1	0
2	0	0,25	0,2	0
3	0	0,5	0,5	0,4
4	0	0	0,2	0,6

Např.:

$$p(3/3) = \frac{p(3,3)}{p_2(3)} = \frac{0,25}{0,5} = 0,5$$

6.1.5. Nezávislost složek náhodného vektoru (X, Y)

Definice 6.1.7.

Náhodná veličina X nezávisí na Y právě tehdy, když jsou podmíněná rozdělení veličiny X stejná jako marginální, pro x:

$p(x/Y=y_0) = p_1(x)$

$f(x/Y=y_0) = f_1(x)$

$F(x/Y=y_0) = F_1(x)$

Poznámka

Je-li náhodná veličina X nezávislá na náhodné veličině Y, pak složka Y je nezávislá na složce X a říkáme, že složky X a Y jsou nezávislé.

Věta 6.1.1.

Je dán náhodný vektor (X,Y). Náhodné veličiny X, Y jsou nezávislé právě tehdy, když platí:

$F(x,y) = F_1(x).F_2(y)$

$p(x,y) = p_1(x).p_2(y)$ ...pro diskrétní náhodný vektor

$f(x,y) = f_1(x).f_2(y)$ ...pro spojitý náhodný vektor

6.2. Číselné charakteristiky náhodného vektoru

Charakteristiky náhodného vektoru (X,Y) slouží k popisu zákona rozdělení pravděpodobnosti náhodného vektoru. Jsou opět konstruovány na základě **počátečního momentu** μ_{kl} nebo **centrálního momentu** ν_{kl}.

Definice 6.2.1. počátečního momentu μ_{kl}

Počáteční momenty $(k+l)$-tého řádu náhodného vektoru (X,Y) jsou střední hodnoty součinu k-tých mocnin složky X a l-tých mocnin složky Y:

$$\mu_{kl} = E\left(X^k.Y^l\right) = \begin{cases} \sum\limits_x \sum\limits_y x^k.y^l.p(x,y) & \text{pro diskrétní náhodnou veličinu} \\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} x^k.y^l.f(x,y)\,dxdy & \text{pro spojitou náhodnou veličinu} \end{cases}$$

Definice 6.2.2. centrálního momentu ν_{kl}

Centrální momenty $(k+l)$-tého řádu náhodného vektoru (X,Y) jsou střední hodnoty součinu k-tých mocnin odchylek složky X od μ_x a l-tých mocnin odchylek složky Y od μ_y:

$$\upsilon_{kl} = \begin{cases} \sum\limits_x \sum\limits_y (x-\mu_x)^k.(y-\mu_y)^l.p(x,y) & \text{pro diskrétní náhodnou veličinu} \\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x-\mu_x)^k.(y-\mu_y)^l.f(x,y)\,dxdy & \text{pro spojitou náhodnou veličinu} \end{cases}$$

6.2.1. Marginální charakteristiky

Tyto charakteristiky popisují vlastnosti marginálních rozdělení jednotlivých složek náhodného vektoru. Popisují tedy odděleně jednotlivé složky náhodného vektoru. Podobně jako u náhodné veličiny popisují polohu, variabilitu, šikmost a špičatost rozdělení. Nejčastěji užívané jsou střední hodnoty a disperze složek:

- **Střední hodnoty náhodných veličin *X* a *Y***

  střední hodnota náhodné veličiny *X*:

$$\mu_{10} = E\left(X^1.Y^0\right) = E(X) = \mu_x = \begin{cases} \sum\limits_i x_i.p_1(x_i) & \text{pro diskrétní náhodnou veličinu} \\ \int\limits_{-\infty}^{\infty} x.f_1(x)\,dx & \text{pro spojitou náhodnou veličinu} \end{cases}$$

  střední hodnota náhodné veličiny *Y*:

$$\mu_{01} = E\left(X^0.Y^1\right) = E(Y) = \mu_y = \begin{cases} \sum\limits_j y_j.p_2(y_j) & \text{pro diskrétní náhodnou veličinu} \\ \int\limits_{-\infty}^{\infty} x.f_2(y)\,dy & \text{pro spojitou náhodnou veličinu} \end{cases}$$

- **Disperze (rozptyl) náhodných veličin *X* a *Y***

  disperze náhodné veličiny *X*:

$$\nu_{20} = D(X) = \sigma_x^2 = \begin{cases} \sum\limits_i \left(x_i - E(X)\right)^2.p_1(x_i) & \text{pro diskrétní náhodnou veličinu} \\ \int\limits_{-\infty}^{\infty} \left(x - E(X)\right)^2.f_1(x)\,dx & \text{pro spojitou náhodnou veličinu} \end{cases}$$

  disperze náhodné veličiny *Y*:

$$\nu_{02} = D(Y) = \sigma_y^2 = \begin{cases} \sum\limits_j \left(y_j - E(Y)\right)^2.p_2(y_j) & \text{pro diskrétní náhodnou veličinu} \\ \int\limits_{-\infty}^{\infty} \left(y - E(Y)\right)^2.f_2(y)\,dy & \text{pro spojitou náhodnou veličinu} \end{cases}$$

6.2.2. Podmíněné charakteristiky

Podmíněné charakteristiky popisují vlastnosti podmíněných rozdělení, tzn., že jde o charakteristiky proměnné X za podmínky, že proměnná Y nabyla určité hodnoty (nebo naopak).

- **Podmíněná střední hodnota *E(X/y)*:**

$$E(X/y)=E(X/Y=y)=\begin{cases}\sum_i x_i . p(x_i/y) & \text{pro diskrétní rozdělení}\\ \int_{-\infty}^{\infty} x.f(x/y)\,dx & \text{pro spojité rozdělení}\end{cases}$$

Protože podmíněná střední hodnota proměnné X závisí na hodnotě veličiny Y, a je tedy její funkcí, nazývá se regresní funkce veličiny X vzhledem k Y.

- **Podmíněná disperze *D*(*X*/*y*)**

$$D(X/y)=E(X/Y=y)=\begin{cases}\sum_i (x_i-E(X/y))^2 .p(x_i/y) & \text{pro diskrétní rozdělení}\\ \int_{-\infty}^{\infty} (x-E(X/y))^2 .f(x/y)\,dx & \text{pro spojité rozdělení}\end{cases}$$

Podmíněná disperze je rovněž závislá na veličině Y. Nazývá se **skedastická funkce** a popisuje, jak se mění rozptyl veličiny X v závislosti na hodnotách proměnné Y. Rozdělení, u kterých je tato funkce konstantní, se nazývají **homoskedastická**.

Poznámka

Vzorce pro E(Y/x), D(Y/x) obdržíme samozřejmě záměnou proměnných X, Y a jejich hodnot x, y.

6.2.3. Charakteristiky popisující vztah mezi proměnnými *X*, *Y*

- **Kovariance cov(*X*, *Y*)**

  Kovariance je střední hodnota součinu odchylek veličin *X* a *Y* od jejich středních hodnot

$$\operatorname{cov}(X,Y)=\nu_{11}=E\left[(X-\mu_x).(Y-\mu_y)\right]=E(X.Y)-E(X).E(Y)=$$

$$=\begin{cases}\sum_i\sum_j x_i.y_j.p(x_i,y_j)-E(X).E(Y) & \text{pro diskrétní náhodný vektor}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x.y.f(x,y)\,dxdy-E(X).E(Y) & \text{pro spojitý náhodný vektor}\end{cases}$$

Platí:

- cov(*X*, *X*) = *D*(*X*)
- cov(*Y*, *Y*) = *D*(*Y*)
- cov(*X*, *Y*) = cov(*Y*, *X*)
- cov(*X*, *Y*) = 0 jsou-li *X* a *Y* nezávislé

- **Koeficient korelace ρ(*X*,*Y*)**

  Koeficient korelace určuje **míru lineární závislosti** náhodných veličin *X* a *Y*

$$\rho(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X).D(Y)}}$$

Vlastnosti:

- $|\rho(X,Y)|\le 1$
- Jestliže $|\rho(X, Y)| = 1$, pak mezi veličinami *X* a *Y* existuje funkční lineární závislost, tzn.: $Y = aX + b$ (*a*, *b* jsou konstanty)
- Jestliže $\rho(X, Y) = 0$, pak veličiny *X* a *Y* jsou nekorelované (nemusí být nezávislé)
- Jestliže $\rho(X, Y) > 0$, pak hovoříme o kladné (přímé) korelaci (obě veličiny současně rostou).

Jestliže $\rho(X, Y) < 0$, pak hovoříme o záporné (nepřímé) korelaci (jedna veličina roste a druhá současně klesá)

- Hodnoty $\rho(X, Y)$ blízké +1 nebo -1 znamenají silnou lineární závislost mezi veličinami X a Y

  Hodnoty $\rho(X, Y)$ blízké 0 znamenají velmi slabou lineární závislost mezi veličinami X a Y.

Řešené úlohy

Příklad 6.2.1. Určete číselné charakteristiky náhodného vektoru (X, Y), který je zadán tabulkou:

$Y \backslash X$	**2**	**3**	**6**
1	0,15	0,20	0,10
3	0,20	0,05	0,30

Řešení: K řešení příkladu můžeme použít např. Excel a vypočítat charakteristiky přesně podle vzorců - viz. tabulka:

	A	B	C	D	E	F
1	$X\backslash Y$	1	3	$p_1(x_i)$	$x_i.p_1(x_i)$	$(x_i-\mu_x)^2.p_1(x_i)$
2	2	0,15	0,2	0,35	0,7	1,197875
3	3	0,2	0,05	0,25	0,75	0,180625
4	6	0,1	0,3	0,4	2,4	1,849
5	$p_2(y_j)$	0,45	0,55	1	**3,85**	**3,2275**
6	$y_j.p_2(y_j)$	0,45	1,65	**2,1**		
7	$(y_j-\mu_y)^2.p_2(y_j)$	0,5445	0,4455	**0,99**		
8	$x_i.y_j.p(x_i,y_j)$	1,5	7,05	**8,55**		
9				**Σ**		

Z tabulky vidíme, že:

$$E(X) = \mu_x = \sum_i x_i p_1(x_i) = \mathbf{3{,}85}$$

$$E(Y) = \mu_y = \sum_j y_j p_2(y_j) = \mathbf{2{,}1}$$

$$D(X) = \sigma_x^2 = \sum_i (x_i - \mu_x)^2 p_1(x_i) = \mathbf{3{,}2275}$$

$$D(Y) = \sigma_y^2 = \sum_j (y_j - \mu_y)^2 p_2(y_j) = \mathbf{0{,}99}$$

$$\operatorname{cov}(X,Y) = \sum_i \sum_j x_i y_j p(x_i, y_j) - E(X).E(Y) = \mathbf{8{,}55} \text{ - } 3{,}85.2{,}1 = 0{,}465$$

$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X).D(Y)}} = \frac{0{,}465}{\sqrt{3{,}2275.0{,}99}} = 0{,}26$$ ... jedná se tedy o slabou lineární závislost

Lze postupovat i jiným způsobem:

Stačí si uvědomit, že pravděpodobnosti v tabulce přesně odpovídají souboru, ve kterém je dvacet uspořádaných dvojic, přičemž např. dvojice (2, 1) se vyskytuje třikrát ($\frac{3}{20} = 0{,}15$), dvojice (2, 3) se vyskytuje čtyřikrát ($\frac{4}{20} = 0{,}2$) ... . Pak stačí přepsat tyto dvojice opět např. do Excelu a využít předdefinovaných funkcí PRŮMĚR, VAR, COVAR, CORREL:

	H	I	J	K	L
1	2	1	**CHARAKTERISTIKY:**		
2	2	1			
3	2	1	**střední hodnoty:**		
4	2	3		**3,85**	(=PRŮMĚR(H1:H20))
5	2	3		**2,1**	(=PRŮMĚR(I1:I20))
6	2	3			
7	2	3		**rozptyly:**	
8	3	1		**3,2275**	(=VAR(H1:H20))
9	3	1		**0,99**	(=VAR(I1:I20))
10	3	1			
11	3	1		**kovariance:**	
12	3	3		**0,465**	(=COVAR(H1:H20;I1:I20))
13	6	1			
14	6	1	**koeficient korelace:**		
15	6	3		**0,260137**	(=CORREL(H1:H20;I1:I20))
16	6	3			
17	6	3			
18	6	3			
19	6	3			
20	6	3			

Tuto úlohu si můžete také otevřít **vyřešenou** v Excelu.

Příklad 6.2.2. Vypočtěte střední hodnotu náhodné veličiny X náhodného vektoru, který je určen hustotou pravděpodobnosti:

$$f(x,y)=\begin{cases}0{,}5.\sin(x+y) & \text{pro } 0\le x\le \frac{\pi}{2},\ 0\le y\le \frac{\pi}{2}\\ 0 & \text{jinde}\end{cases}$$

Řešení:

$$E(X)=\int_{-\infty}^{\infty} x.f_1(x)\,dx,\ \text{kde } f_1(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$$

$$E(X)=\frac{1}{2}\int_0^{\frac{\pi}{2}}dx\int_0^{\frac{\pi}{2}}x.\sin(x+y)\,dy=\frac{1}{2}\int_0^{\frac{\pi}{2}}dx\Big[-x.\cos(x+y)\Big]_0^{\frac{\pi}{2}}=$$

$$=\frac{1}{2}\int_0^{\frac{\pi}{2}}x\left[-\cos\left(x+\frac{\pi}{2}\right)+\cos x\right]dx=per\ partes=$$

$$=\left|\begin{matrix}u=x & v'=-\cos\left(x+\frac{\pi}{2}\right)+\cos x\\ u'=1 & v=-\sin\left(x+\frac{\pi}{2}\right)+\sin x\end{matrix}\right|=$$

$$=\frac{1}{2}\left[x.\left[-\sin\left(x+\frac{\pi}{2}\right)+\sin x\right]\right]_0^{\frac{\pi}{2}}+\frac{1}{2}\int_0^{\frac{\pi}{2}}\left[\sin\left(x+\frac{\pi}{2}\right)-\sin x\right]dx=$$

$$=\frac{\pi}{4}+\frac{1}{2}\left[-\cos\left(x+\frac{\pi}{2}\right)+\cos x\right]_0^{\frac{\pi}{2}}=\frac{\pi}{4}+\frac{1}{2}(1-1)=\frac{\pi}{4}$$

Podobným způsobem by se daly vypočítat i zbylé číselné charakteristiky: disperze, kovariance a koeficient korelace.

Úlohy k samostatnému řešení

6.1. Náhodný vektor (X,Y) má pravděpodobnostní funkci zadanou tabulkou:

$X \backslash Y$	1	2	3
-1	0,15	0,05	0,10
0	0,10	0,10	0,15
1	0,05	0,10	0,20

Určete:

a) $P(X = 0, Y = 3)$

b) $P(X < 0{,}5, Y < 2{,}5)$

c) $P(X > 0, Y > 2{,}5)$

d) marginální rozdělení

e) distribuční funkci

6.2. Náhodný vektor je dán pravděpodobnostní funkcí:

$X \backslash Y$	0	1	2
2	0,15	0,2	0,3
3	0,05	0,2	?

Doplňte chybějící hodnotu a určete marginální pravděpodobnostní funkci a sdruženou distribuční funkci.

6.3. V sérii výrobků měříme jejich délku s přesností 0,5 mm a šířku s přesností 0,2 mm. Označme jako náhodnou veličinu X chybu, které se dopustíme při měření délky a Y při měření šířky. Za předpokladu rovnoměrného rozdělení určete pravděpodobnost, že délka bude měřena s max. chybou 0,2 mm a současně šířka s max. chybou 0,1 mm.

6.4. Určete střední hodnoty, disperze, kovarianci a koeficient korelace náhodného vektoru, který je popsán pravděpodobnostní funkcí:

a)

$X\backslash Y$	0	1	2	3
0	0,008	0,036	0,054	0,027
1	0,060	0,180	0,135	0
2	0,150	0,225	0	0
3	0,125	0	0	0

b}

$X\backslash Y$	1	2	3	4
3	0,01	0,02	0,03	0,25
5	0,04	0,16	0,18	0,05
7	0,12	0,07	0,06	0,01

c)

$X\backslash Y$	-2	2	6
2	0,6	0	0
4	0	0,2	0
6	0	0	0,2

6.5. Pro náhodný vektor daný následující tabulkou vypočtěte koeficient korelace

$X\backslash Y$	1	0
1	0,005	0,01
0	0,02	0,965

Výsledky úloh k samostatnému řešení

6.1. *a*) 0,15

b) 0,4

c) 0,2

6.2. ? = 0,1

6.3. 0,2

6.4. *a*) 1,5; 0,9; 0,75; 0,63; -0,45; -0,654

b) 4,9; 2,72; 2,27; 1,1616; -1,048; -0,64539

c) 3,2; 0,4; 2,56; 10,24; 5,12; 1

6.5. 0,2445

7. STATISTICKÝ SOUBOR S JEDNÍM ARGUMENTEM

Průvodce studiem

Předchozí kapitoly byly věnovány pravděpodobnosti a tomu, co s tímto pojmem souvisí. Nyní znalosti z počtu pravděpodobnosti aplikujeme ve statistice.

Předpokládané znalosti

Pojmy z předchozích kapitol.

Cíle

Cílem této kapitoly je zavést a objasnit pojem statistika, seznámit se základní statistickou terminologií a definovat charakteristiky statistického souboru s jedním argumentem.

Výklad

7.1. Úvod do statistiky

Několik citátů na úvod:

Nevěřím jiné statistice, než té, kterou jsem osobně zfalšoval.

Winston Churchill

Statistika je obzvláště rafinovaná forma lži.

???

S pomocí statistiky je jednoduché lhát. Bez ní je ale těžké říci pravdu.

Andrejs Dunkels

Už z těchto vět je patrné, že statistika měla a má poněkud pošramocenou pověst vědy, která má často vytvářet pouze jakousi iluzi pravdy a jejíž přímým úkolem je někdy skutečnost úmyslně mást (na obranu statistiky i W. Churchilla nutno poznamenat, že v případě prvního citátu se pravděpodobně jedná o podvrh, fámu o tomto údajném Churchillově výroku rozšířil německý ministr propagandy Joseph Goebbels).

Jak jednoduché je ze správných statistických údajů vyvodit nesmyslné závěry, můžeme dokumentovat na následujícím příkladě: Je statisticky dokázáno, že každé čtvrté dítě, které se narodí, je Číňan. Znamená to však něco při plánování počtu dětí pro průměrnou českou rodinu? Většina čtenářů asi tuší, že nikoliv. Jsme však schopni takový rozpor vždy odhalit?

Abychom se tedy vyvarovali nesprávných úsudků vyplývajících z neznalosti, je vhodné se seznámit se základy matematické statistiky a s jejími možnostmi.

Nejčastější aplikace počtu pravděpodobnosti směřují do oblasti statistiky. Její nejrozšířenější část, tzv. **matematická statistika,** se zabývá metodami získávání, zpracování a vyhodnocování hromadných dat (tzn. údajů o vlastnostech velkého počtu jedinců - osob, věcí či jevů).

Podle použitých metod práce dělíme matematickou statistiku na

- **deskriptivní, popisnou statistiku** - zabývá se efektivním získáváním ukazatelů, které poskytují obraz zkoumaného jevu;
- **statistickou indukci** (matematickou statistiku v užším smyslu) - řeší problémy zobecňování výsledků získaných popisem statistického souboru.

7.2. Statistický soubor s jedním argumentem - základní pojmy

Množinu všech předmětů pozorování (osob, věcí, jevů apod.) shromážděných na základě toho, že mají společné vlastnosti, nazýváme **statistickým souborem**. Jednotlivé prvky této množiny se nazývají prvky (elementy) statistického souboru nebo též statistické jednotky. Počet všech prvků statistického souboru se nazývá **rozsah souboru *N*.**

Soubor, který je předmětem zkoumání, se nazývá základní soubor. Často nelze nebo není účelné provést zkoumání všech statistických jednotek tohoto základního souboru. Základní soubor pak zkoumáme pomocí statistických jednotek, které z něj byly určitým způsobem vybrány a které tvoří takzvaný výběrový soubor.

Poznámka

Například: Při zjišťování výšky studentů ve studijní skupině je statistickým souborem množina studentů dané skupiny. Jejich společnou vlastností je, že jsou studenty například studijní skupiny JB007 Vysoké školy báňské, a že budeme zkoumat jejich výšku. Statistickou jednotkou je student dané skupiny. Rozsahem souboru je počet studentů dané skupiny, například 21. Statistickým souborem může být také množina všech studentů této školy.

Vlastnosti statistických souborů, které jsou předmětem statistického zkoumání, sleduje statistika prostřednictvím vlastností statistických jednotek daného souboru, které postihuje statistickými znaky. **Statistický znak** je vyjádřením určité vlastnosti statistických jednotek (prvků množin) sledovaného statistického souboru; slouží k charakterizování sledovaného hromadného jevu-vlastnosti daného statistického souboru. Znak (argument) souboru se zpravidla značí x. Jednotlivé údaje znaku se nazývají hodnoty znaku, značí se x_1, x_2, x_N, kde N je rozsah souboru.

Poznámka

Například: Například při určování výšky studentů dané studijní skupiny je statistickým znakem výška studentů, hodnotou znaku je číselně vyjádřená příslušná výška studenta, např.182 cm.

Hodnoty znaku mohou být vyjádřeny buď čísly nebo jiným způsobem (zpravidla slovním popisem). V prvním případě mluvíme o znacích kvantitativních, např. tělesná výška, tělesná hmotnost, počet obyvatel měst, atp.. V druhém případě mluvíme o znacích kvalitativních, které se mohou vyskytovat ve dvou druzích (znaky alternativní, např. muž-žena, voják-nevoják, prospěl-neprospěl) nebo ve více druzích (např. povolání, národnost, náboženství, atp.).

Další pojmy

Když $x_m = \min\limits_i \; x_i$ a $x_M = \max\limits_i \; x_i$, pak interval $\langle x_m, x_M \rangle$ je **variační obor** argumentu X.

Hodnota $R = x_M - x_m$ je **variační rozpětí** argumentu X.

Jestliže se hodnota x_i vyskytne v souboru f_i-krát, je f_i **absolutní četnost hodnoty x_i**.

Hodnoty x_i seřazené podle velikosti a jejich absolutní četnosti f_i tvoří **variační řadu** (statistickou řadu).

Hodnota $\varphi_i = \dfrac{f_i}{N}$ (N je rozsah souboru) je **relativní četnost** hodnoty x_i.

Hodnota $F_i = \sum_{k=1}^{i} f_k$ je **kumulativní četnost** do x_i.

Hodnota $\Phi_i = \frac{F_i}{N}$ je **relativní kumulativní četnost** do x_i.

Řešené úlohy

Příklad 7.2.1. Určete relativní, kumulativní a relativní kumulativní četnosti variační řady

x_i	0	1	2	3	4
f_i	7	44	56	30	12

Řešení:

$$N = \sum_{i=1}^{5} f_i = 149$$

Všechny četnosti vypočteme z výše uvedených vzorců:

x_i	0	1	2	3	4	Σ
f_i	7	44	56	30	12	149
ϕ_i	0,047	0,295	0,376	0,201	0,081	1
F_i	7	51	107	137	149	
Φ_i	0,047	0,342	0,718	0,919	1	

7.3. Charakteristiky statistického souboru s jedním argumentem

Charakteristiky statistických souborů se definují analogicky jako charakteristiky náhodné proměnné ***X***, jíž u statistických souborů je uvažovaný argument. Úlohu pravděpodobnosti hrají zde relativní četnosti (ve shodě se statistickou definicí pravděpodobnosti) a funkce *φ(x)* a *Φ(x)* lze považovat za empirické pravděpodobnostní funkce variační řady s analogickými vlastnostmi, jaké mají funkce rozložení pravděpodobnosti náhodné veličiny.

Mezi nejdůležitější charakteristiky patří **charakteristiky polohy**, střední hodnota, modus, medián a kvantily.

Definice 7.3.1.

Empirická střední hodnota je

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i \ .$$

Modus statistického souboru $Mo(x)$

je ta hodnota argumentu X, která má největší absolutní četnost.

Medián statistického souboru $Me(x)$

je ta hodnota argumentu X, která rozděluje soubor uspořádaný na dvě části o stejném počtu prvků. Má-li soubor sudý počet prvků, považuje se za medián průměrná hodnota prostředních dvou.

Empirický *p*-kvantil

je taková hodnota x_p, pro kterou platí, že $100p$ procent prvků souboru je nanejvýš rovných x_p.

Nejčastěji používanými kvantily jsou **kvartily**, **decily** a **percentily**. Definujte je. A co je z hlediska kvantilů vlastně medián?

Druhou skupinu charakteristik jsou **charakteristiky variability**, empirický rozptyl (disperze), směrodatná (standardní) odchylka, průměrná odchylka a variační koeficient. Většina z nich je přímou analogií příslušných teoretických ukazatelů.

Definice 7.3.2.

Empirický rozptyl (empirická disperze) je dán vztahem

$$s_x^2 = D\ \ x\ \ = \frac{1}{N}\sum_i f_i\ \ x_i - \bar{x}\ \ ^2$$

Empirická směrodatná (standardní) odchylka je

$$s_x = \sqrt{D\ \ x}$$

Průměrná odchylka je určena vztahem

$$\bar{d} = \frac{1}{N}\sum_i f_i . \left| x_i - \bar{x} \right|$$

Variační koeficient je dán vztahem

$v = \frac{s_x}{x}$ (často se udává v procentech).

Poznámky

Základní vlastnosti směrodatné odchylky:

- *směrodatná odchylka měří rozptýlenost kolem průměru*

  $s = 0$ pouze v případech, kdy se všechna data rovnají stejné hodnotě, jinak $s > 0$

- *stejně jako průměr je i směrodatná odchylka silně ovlivněna extrémními hodnotami, i jedna nebo dvě odlehlé hodnoty ji silně zvětšují*

- *je-li rozdělení dat silně zešikmené (zjistíme pomocí koeficientu šikmosti), směrodatná odchylka neposkytuje dobrou informaci o rozptýlenosti dat - v těchto případech používáme kvantilové charakteristiky - viz. dále*

Variační koeficient používáme, jestliže chceme posoudit relativní velikost rozptýlenosti dat vzhledem k průměru. Počítáme ho, když chceme porovnat rozptýlenost dat skupin měření stejné proměnné s různým průměrem, nebo v případech, kdy se mění velikost směrodatné odchylky tak, že je přímo závislá na úrovni měřené proměnné.

Důležitou roli opět i ve statistice hrají momentové charakteristiky. Uveďme jen jejich definice značené latinskými ekvivalenty řeckých označení z počtu pravděpodobnosti.

Definice 7.3.3.

Počáteční empirický moment *k*-tého řádu

$$m_k = \frac{1}{N} \sum_{i=1}^{n} f_i x_i^k$$

Centrální empirický moment *k*-tého řádu

$$n_k = \frac{1}{N} \sum_{i=1}^{n} f_i \left(x_i - \bar{x} \right)^k$$

Normovaný empirický moment *k*-tého řádu

$$n_k = \frac{n_k}{s_x^k}$$

Samozřejmě platí analogické vztahy pro výpočty momentů centrálních z počátečních:

$n_2 = m_2 - m_1^2$

$n_3 = m_3 - 3m_2m_1 + 2m_1^3$

$n_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$

Normované momenty použijeme i tady jako ukazatele šikmosti a špičatosti:

Definice 7.3.4.

Empirický koeficient šikmosti

$$A = n_3 = \frac{n_3}{s^3}$$

Empirický exces

$$e = n_4 - 3 = \frac{n_4}{s^4} - 3$$

Řešené úlohy

Příklad 7.3.1. Vypočtěte empirické charakteristiky, modus a kvartily variační řady:

x_i	0	1	2	3	4
f_i	7	44	51	30	12

Řešení: Ukážeme tři způsoby výpočtu v Excelu:

Nejdříve charakteristiky vypočteme přesně podle vzorců, které jsme uvedli:

	A	B	C	D	E	F
1	x_i	f_i	m_1	n_2	n_3	n_4
2	0	7	0	27,22762	-53,6989	105,9062
3	1	44	44	41,58951	-40,4342	39,31107
4	2	51	102	0,039352	0,001093	3,04E-05
5	3	30	90	31,68981	32,57009	33,47481
6	4	12	48	49,34259	100,0558	202,891
7	**Σ**	**144**	**1,972222**	**1,040895**	**0,267318**	**2,649882**

Z tabulka snadno dopočteme číselné charakteristiky:

Střední hodnota:

$$\bar{x} = m_1 = \frac{1}{N}.\sum_{i=1}^{5} f_i.x_i = 1{,}97\overline{2}$$

Rozptyl:

$$s^2 = n_2 = \frac{1}{N}.\sum_{i=1}^{5} f_i.\left(x_i - \bar{x}\right)^2 \approx 1{,}041$$

Směrodatná odchylka:

$$s_x = \sqrt{1{,}041} \approx 1{,}020$$

Koeficient šikmosti:

$$A_x = n = \frac{n_3}{s^3} = \frac{\frac{1}{N}.\sum_{i=1}^{5} f_i.\left(x_i - \bar{x}\right)^3}{s^3} = \frac{0{,}267}{1{,}02^3} \approx 0{,}252$$

Exces:

$$e = n_4 = \frac{n_4}{s^4} - 3 = \frac{2{,}65}{1{,}02^4} - 3 \approx -0{,}554$$

Modus: největší absolutní četnost má hodnota 2, takže:

$Mo(x) = 2$

Při výpočtu kvartilů určíme nejprve jejich pořadí podle vzorce:

$z_p = N.p + 0{,}5$, tedy:

$z_{0,25} = 144.0{,}25 + 0{,}5 = 36{,}5$

$z_{0,5} = 144.0{,}5 + 0{,}5 = 72{,}5$

$z_{0,75} = 144.0{,}75 + 0{,}5 = 108{,}5$

Z výpočtu pořadí vidíme, že 1.kvartil se vypočte jako aritmetický průměr hodnot 36 a 37 prvku - z tabulky je zřejmé, že obě jsou rovny 1, tzn.

$x_{0,25} = 1$, obdobně

$x_{0,5} = 2$ (medián)

$x_{0,75} = 3$

Druhá možnost je použití předdefinovaných funkcí v Excelu:

	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	VÝPOČET POMOCÍ PŘEDDEFINOVANÝCH FUNKCÍ												
2	DATA:												
3	0	0	0	0	0	0	0	1	1	1	1	1	
4	1	1	1	1	1	1	1	1	1	1	1	1	
5	1	1	1	1	1	1	1	1	1	1	1	1	
6	1	1	1	1	1	1	1	1	1	1	1	1	
7	1	1	1	2	2	2	2	2	2	2	2	2	
8	2	2	2	2	2	2	2	2	2	2	2	2	
9	2	2	2	2	2	2	2	2	2	2	2	2	
10	2	2	2	2	2	2	2	2	2	2	2	2	
11	2	2	2	2	2	2	3	3	3	3	3	3	
12	3	3	3	3	3	3	3	3	3	3	3	3	
13	3	3	3	3	3	3	3	3	3	3	3	3	
14	4	4	4	4	4	4	4	4	4	4	4	4	
15													
16			střední hodnota						modus				
17	1,972222	(=PRŮMĚR(J3:U14))						2		(=MODE(J3:U14))			
18			rozptyl						1.kvartil				
19	1,040895	(=VAR(J3:U14))						1		(=QUARTIL(J3:U14;1)			
20			směrodatná odchylka						2.kvartil				
21	1,020243	(=SMODCH(J3:U14))						2		(=QUARTIL(J3:U14;2))			
22			koeficient šikmosti						3.kvartil				
23	0,254378	(=SKEW(J3:U14))						3		(=QUARTIL(J3:U14;3))			
24			exces										
25	-0,53113	(=KURT(J3:U14))											

Pro pokročilé uživatele Excelu bude možná nejvhodnější třetí možnost, jak vyřešit tuto úlohu. Použijeme doplňkový nástroj Excelu, který se nazývá Analýza dat. Pokud v menu Excelu v nabídce Nástroje nenajdete tento nástroj, je nutné ho doinstalovat. Tento úkon je velmi jednoduchý. V nabídce Nástroje klepněte na příkaz Doplňky. V seznamu Doplňky k dispozici zaškrtněte políčko u položky Analytické nástroje a klepněte na tlačítko **OK**. Po instalaci by mělo být možné doplněk spustit z nabídky Nástroje.

Chceme-li vypočítat příslušné charakteristiky, data umístíme do jednoho sloupce (řádku) a v dialogovém okně Analýza dat klepneme na analytický nástroj Popisná statistika a nastavíme požadované možnosti analýzy.

Výstup pak v našem příkladě vypadá takto:

Sloupec1	
Stř. hodnota	1,972222
Chyba stř. hodnoty	0,085317
Medián	2
Modus	2
Směr. odchylka	1,023804
Rozptyl výběru	1,048174
Špičatost	-0,53113
Šikmost	0,254378
Rozdíl max-min	4
Minimum	0
Maximum	4
Součet	284
Počet	144
Největší (1)	4
Nejmenší (1)	0

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

7.4. Zpracování rozsáhlého statistického souboru

Obsahuje-li statistický soubor velký počet různých hodnot argumentu X, sdružujeme hodnoty argumentu do intervalů zvaných třídy. Obvykle volíme konstantní šířku třídy.

Hranice tříd je nutno volit tak, aby každý prvek statistického souboru bylo možné zařadit právě do jedné třídy.

Počet tříd volíme podle účelu zkoumání, obvykle 5-20 tříd. Přesné pravidlo pro výpočet počtu tříd neexistuje. Uvedeme alespoň některé doporučované možnosti:

- pro šířku třídy h by mělo přibližně platit

  $h \approx 0{,}08 \cdot x_{\max} - x_{\min}$,

- počet tříd n by měl být

  $n \approx 1 + 3{,}3 \cdot \log N$ nebo

  $n \leq 5 \cdot \log N$ nebo

  $n \approx \sqrt{N}$,

- pro $30 \leq N < 100$ volíme 7-10 tříd,

  pro $100 \leq N < 500$ volíme nejvýše 15 tříd,

  pro $N \geq 500$ volíme nejvýše 20 tříd.

Při zpracování statistického souboru nahradíme všechny hodnoty v dané třídě jedinou hodnotou, tzv. **třídním znakem**, kterým je **aritmetický průměr** obou mezí třídy. Třídní znak zastupuje všechny hodnoty, které do této třídy patří. Počet hodnot ve třídě je **třídní četnost**.

Po rozdělení souboru do tříd už nepočítáme s jednotlivými hodnotami, ale s třídami, třídními znaky a třídními četnostmi. Rozdělením variačního oboru na třídy a shrnutím všech hodnot argumentu v každé třídě do **třídního znaku** se dopouštíme při výpočtu centrálních momentů systematických chyb. Anglický statistik **W. F. Shepard** odvodil v r. 1897 korekce, jimiž lze tyto chyby korigovat.

Značí-li ***h*** šířku tříd, jsou opravené momenty dány vzorci:

Shepardovy korekce

$n_1 = n_1, \qquad n_3 = n_3$ (liché momenty se neopravují)

$$n_2 = n_2 - \frac{h^2}{12}, \qquad n_4 = n_4 - n_2.\frac{h^2}{2} + \frac{7}{240}.h^4$$

Modus se u rozsáhlého statistického souboru, který je rozdělen do tříd, vypočte interpolací:

$$Mo\ x = x_j - \frac{h}{2}.\frac{f_{j+1} - f_{j-1}}{f_{j+1} + f_{j-1} - 2f_j}$$

x_j ... střed j-té třídy s největší absolutní četností f_j

h ... šířka třídy

Kvantily se v tomto případě určí opět interpolací:

$$x_p = x_j - \frac{h}{2} + N.p - F_{j-1}\ .\frac{h}{f_j}$$

j ... pořadí třídy, do níž je zařazen ($N.p$)-tý prvek uspořádaného souboru

x_j ... střed j-té třídy

F_{j-1} ... kumulativní absolutní četnost (j - 1)-vé třídy

f_j ... absolutní četnost j-té třídy

Řešené úlohy

Příklad 7.4.1. Na jednom nejmenovaném pracovišti byly při zjišťování IQ naměřeny následující hodnoty:

68, 71, 71, 78, 82, 82, 87, 91, 92, 92, 95, 97, 102, 102, 102, 103, 105, 105, 109, 110, 111, 111, 111, 112, 112, 114, 114, 114, 115, 116, 118, 119, 121, 122, 122, 124, 126, 131, 133, 137.

Rozdělte tyto hodnoty do osmi tříd a určete empirické charakteristiky, modus a kvartily.

Řešení:

$x_{max} - x_{min} = 137 - 68 = 69$

Vypočteme šířku třídy:

$$h = \frac{69}{8} = 8{,}625 \doteq 9$$

Když ale nyní vynásobím 9.8 = 72, to je o tři více než původně vypočtené variační rozpětí. Dolní hranici 1.třídy proto zvolím o 1,5 menší, než je x_{min}, tedy 66,5.

K výpočtu empirických charakteristik je vhodné použít např. Excel - viz. tabulka:

	A	B	C	E	F	G	H
1	třídy	třídní znak x_i	třídní četnost f_i	m_1	n_2	n_3	n_4
2	66,5 - 75,5	71	3	5,325	90,04669	-3120,12	108112,1
3	75,5 - 84,5	80	3	6	49,34419	-1265,68	32464,65
4	84,5 - 93,5	89	4	8,9	27,72225	-461,575	7685,231
5	93,5 - 102,5	98	5	12,25	7,315312	-55,9621	428,1104
6	102,5 - 111,5	107	8	21,4	0,3645	0,492075	0,664301
7	111,5 - 120,5	116	9	26,1	24,10256	249,4615	2581,927
8	120,5 - 129,5	125	5	15,625	46,80281	905,6344	17524,03
9	129,5 - 138,5	134	3	10,05	60,27919	1708,915	48447,74
10		**Σ**	**40**	**105,65**	**305,9775**	**-2038,83**	**217244,4**

Z hodnot v tabulce pak snadno vypočteme hledané charakteristiky:

Empirická střední hodnota:

$$\bar{x} = m_1 = \frac{1}{N}.\sum_{i=1}^{8} f_i.x_i = 105{,}65$$

Empirická disperze:

$$s^2 = n_2 = n_2 - \frac{h^2}{12} = \frac{1}{N}.\sum_{i=1}^{8} f_i.\left(x_i - \bar{x}\right)^2 - \frac{8^2}{12} =$$

$$= 305{,}9775 - 5{,}33 \doteq 300{,}64$$

Empirická směrodatná odchylka:

$$s_x = \sqrt{300{,}64} \doteq 17{,}34$$

Empirický koeficient šikmosti:

$$A_x = n = \frac{n_3}{s^3} = \frac{\frac{1}{N}.\sum_{i=1}^{8} f_i.\left(x_i - \bar{x}\right)^3}{s^3} = \frac{-2038{,}83}{17{,}34^3} \doteq -0{,}39$$

Empirický exces:

$$e = n_4 - 3 = \frac{n_4}{s^4} - 3 = \frac{n_4 - n_2 . \frac{h^2}{2} + \frac{7}{240} . h^4}{s^4} - 3 =$$

$$= \frac{217244,4 - 305,9775 . \frac{64}{2} + \frac{7}{240} . 8^4}{17,34^4} - 3 \approx -0,704$$

Modus:

$$Mo \quad x \quad = x_j - \frac{h}{2} . \frac{f_{j+1} - f_{j-1}}{f_{j+1} + f_{j-1} - 2 f_j} = 116 - \frac{9}{2} . \frac{5-8}{5+8-2.9} = 113,3$$

K výpočtu kvartilů budeme potřebovat ještě tabulku kumulativních třídních četností F_i:

	A	B	C	D
1	třídy	třídní znak x_i	třídní četnost f_i	kumulativní tř. četnost F_i
2	66,5 - 75,5	71	3	3
3	75,5 - 84,5	80	3	6
4	84,5 - 93,5	89	4	10
5	93,5 - 102,5	98	5	15
6	102,5 - 111,5	107	8	23
7	111,5 - 120,5	116	9	32
8	120,5 - 129,5	125	5	37
9	129,5 - 138,5	134	3	40

1.kvartil:

$N.p = 40.0,25 = 10$

10-tý prvek leží ve třetí třídě, tudíž $j = 3$

$$x_{0,25} = x_3 - \frac{h}{2} + \quad N.p - F_{3-1} \quad . \frac{h}{f_3} = 89 - \frac{9}{2} + \quad 10 - 6 \quad . \frac{9}{4} = 93,5$$

2.kvartil (medián):

$N.p = 40.0,5 = 20$

20-tý prvek leží v páté třídě, tudíž $j = 5$

$$x_{0,5} = x_5 - \frac{h}{2} + \quad N.p - F_{5-1} \quad . \frac{h}{f_5} = 107 - \frac{9}{2} + \quad 20 - 15 \quad . \frac{9}{8} = 108,125$$

3.kvartil:

$N.p$ = 40.0,75 = 30

30-tý prvek leží v šesté třídě, tudíž j = 6

$$x_{0,75} = x_6 - \frac{h}{2} + \ N.p - F_{6-1}\ .\frac{h}{f_6} = 116 - \frac{9}{2} + \ 30 - 23\ .\frac{9}{9} = 118,5$$

Pro srovnání ještě uvedeme hodnoty charakteristik, vypočtené (opět v Excelu) bez rozdělení do tříd:

J	K	L	M	N	O
VÝPOČET CHARAKTERISTIK BEZ ROZDĚLENÍ DO TŘÍD					
DATA:					
68	71	71	78	82	
82	87	91	92	92	
95	97	102	102	102	
103	105	105	109	110	
111	111	111	112	112	
114	114	114	115	116	
118	119	121	122	122	
124	126	131	133	137	
	střední hodnota:				modus:
	105,675				102
	disperze:				kvartily:
	292,469375				94,25
směrodatná odchylka:					110,5
	17,10173602				116,5
koeficient šikmosti:					
	-0,510326477				
	exces:				
	-0,292289464				

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Poznámka

Způsob zpracování statistických dat závisí na tom, jak jsou vstupní data zadána (netříděný soubor individuálních hodnot, tříděný soubor - četnostní tabulka), jak velký je rozsah souboru, zda je ke zpracování možno použít výpočetní techniky. Tvar výpočetních tabulek, *které je třeba při výpočtech vytvořit, je dost individuální. I při "ručním" zpracování dat je však možno doporučit metody práce, jaké jsou běžné v tabulkových kalkulátorech, např. v excelu.*

Pro práci se statickými soubory si zopakujte základní výpočetní postupy v excelu. Vyhledejte v nabídce vestavěných funkcí, které z nich odpovídají funkcím, které jsme uváděli jako charakteristiky statistického souboru (kategorie statistických funkcí, ale k některým triviálním výpočtům použijeme i některé funkce matematické).

Ještě jeden citát na závěr:

Statistik je ten, kdo s hlavou v rozpálené troubě a s nohama v nádobě s ledem na dotaz, jak se cítí, odpoví: "V průměru se cítím dobře."

anonym

Úlohy k samostatnému řešení

7.1. Při zjišťování IQ na jednom nejmenovaném pracovišti byly naměřeny tyto hodnoty:
68, 71, 71, 78, 82, 82, 87, 91, 92, 92, 95, 97, 102, 102, 102, 103, 105, 105, 109, 110, 111, 111, 111, 112, 112, 114, 114, 114, 115, 116, 118, 119, 121, 122, 122, 124, 126, 131, 133, 137.
Rozdělte hodnoty do 8 tříd a určete empirické charakteristiky, modus a kvartily.

7.2. Určete medián a střední hodnotu měsíční spotřeby elektrické energie (kWh) v bytech z následujících údajů:
169, 108, 26, 43, 114, 68, 35, 183, 103, 266, 74, 205, 62, 230, 85, 487, 120, 148, 91, 18, 58, 96, 295, 42, 137

7.3. Student se připravuje na zkoušku. Zjistil, že musí nastudovat průměrně 20 stran denně. První polovinu knihy studoval s rychlostí 10 stran denně. Stihne studium celé látky v určeném termínu, bude-li druhou polovinu studovat rychlostí 30 stran denně? Určete průměrný počet stran, které denně nastudoval.

7.4. Zkoušky životnosti žárovek daly následující výsledky (v hodinách):
606, 1249, 267, 44, 510, 340, 109, 1957, 463, 801, 1082, 169, 233, 1734, 1458, 80, 1023, 2736, 917, 459.
Určete střední dobu životnosti žárovek a jejich disperzi.

7.5. Sledovaný statistický znak nabyl těchto hodnot:
60, 80, 80, 100, 100, 100, 100, 120, 120, 150, 150, 160, 180, 200, 200, 200, 200, 200, 220, 250, 250, 250, 280, 300, 300, 300, 300, 350, 350, 360, 380, 400, 400, 400, 400, 420, 450, 500, 500, 550
Určete střední hodnotu a disperzi tohoto souboru. Určete tyto charakteristiky také pro tento soubor roztříděný do tříd:

a) 0-99, 100-199, ...

b) 55-155, 155-255, ...

a porovnejte výsledky obou třídění.

7.6. Určete momentové charakteristiky, modus a kvartily následujícího, do tříd rozděleného, souboru. Použijte Sheppardových korekcí.

x_i	390	410	430	450	470	490	510	530	550	570
fi	7	10	14	22	25	12	3	3	2	2

Výsledky úloh k samostatnému řešení

7.2. $x_{0,5}$ = 103kWh, x = 130,52kWh

7.3. ne, 15

7.4. $x = 811,85$; $s_x^2 = 493407$

7.5. $x = 260,25$; $s^2 = 17342$; $x_1 = 282,5$; $s_1^2 = 19194$; $x_2 = 257,5$; $s_2^2 = 16494$

7.6. $x = 457,4$; $s_x^2 = 1459,9$; $s_x = 38,2$; $A_x = 0,536$; $e = 0,575$;
$x_{0,25} = 431,4$; $x_{0,5} = 457,3$; $x_{0,75} = 477,6$; $Mo(x) = 463,75$

8. STATISTICKÝ SOUBOR SE DVĚMA ARGUMENTY

Průvodce studiem

Využijeme znalostí z předchozí kapitoly, která pojednávala o statistickém souboru s jedním argumentem a rozšíříme je.

Předpokládané znalosti

Pojmy z předchozích kapitol.

Cíle

Cílem této kapitoly je seznámit se statistickým souborem se dvěma argumenty a jeho charakteristikami.

Výklad

8.1. Statistický soubor se dvěma argumenty

Vezměme v úvahu statistický soubor rozsahu N. U každého prvku sledujme hodnoty dvou statistických znaků, dvou argumentů X, Y. Tak vznikne **statistický soubor se dvěma argumenty**.Statistické znaky sledované současně na každém statistickém prvku (nositeli) mohou být diskrétní nebo spojité. Budou nás pochopitelně zajímat hodnoty každého znaku samostatně, ale i jak jsou rozloženy různé kombinace obou znaků. Tak např. u souboru lidí nás mohou zajímat dva antropologické znaky, tělesná výška a tělesná váha. Výrobce oděvů nezajímá jen rozložení výšek, ale simultánně i vah, neboť rozměry oblečení musí být úměrně vyráběny i pro všechny možné existující kombinace hodnot těchto znaků.

Zadání dvojrozměrné diskrétní náhodné veličiny je možno provést v podstatě dvojím způsobem, a to buď pomocí tzv. četnostní plošné tabulky se dvěma vstupy x_i a y_j nebo lineární tabulkou dvojic (x_i, y_i), kde x a y jsou jednotlivé realizace náhodných veličin X a Y. Počet výskytů konkrétní dvojice (x_i, y_j) se nazývá **četnost** (absolutní) $f_{i,j}$.

Podíl $\dfrac{f_{i,j}}{N} = \varphi_{i,j}$ je pak **četnost relativní** . Druhý zápis vyjadřuje funkční hodnotu empirické funkce rozložení pravděpodobnosti dvojrozměrné náhodné veličiny, jejíž realizaci statistický soubor představuje.

Zadání plošnou tabulkou je běžnější pro rozsáhlejší soubory dat, u nichž opakování výskytu jednotlivých dvojic je častější.

Takto např. vypadá zadání v excelu:

	A	B	C	D	E	F	G	H	I
1									
2		**x \ y**	**20**	**30**	**40**	**50**	**60**	**70**	**80**
3		**250**	19	5					
4		**350**	23	116	11				
5		**450**	1	41	98	9			
6		**550**		4	32	65	7		
7		**650**		1	4	21	46	3	
8		**750**			1	2	11	13	1
9		**850**					1	3	2
10									
11									

Zaveďme následující označení:

$X \setminus Y$	y_1	y_2	…	y_k	…	y_n	$\sum$
x_1	f_{11}	f_{12}	…	f_{1k}	…	f_{1n}	M_1
…	…	…	…	…	…	…	…
x_i	f_{i1}	f_{i2}	…	f_{ik}	…	f_{in}	M_i
…	…	…	…	…	…	…	…
x_m	f_{m1}	f_{m2}	…	f_{mk}	…	f_{mn}	M_m
$\sum$	N_1	N_2	…	N_k	…	N_n	N

Pro okrajové sumy platí:

$$M_i = \sum_{k=1}^{n} f_{ik} \quad , \; N_k = \sum_{i=1}^{m} f_{ik} \quad \text{... marginální četnosti hodnot } x_i \text{ a } y_j$$

$$\text{a celkem je: } \sum_{i=1}^{m} \sum_{k=1}^{n} f_{ik} = \sum_{k=1}^{n} N_k = \sum_{i=1}^{m} M_i = N$$

Pro posouzení vlastností náhodné dvojrozměrné veličiny se používají opět momentové charakteristiky analogické veličinám s jedním argumentem.

Tak **počáteční moment (*r* + *s*)-tého stupně** je definován jako číslo

$$m_{r,s} = \frac{1}{N}\sum_i \sum_j x_i^r y_j^s f_{i,j} = \sum_i \sum_j x_i^r y_j^s \varphi_{i,j} \ ,$$

když sčítání proběhne přes všechny hodnoty *i* a *j* jako ve výše uvedené četnostní tabulce.

Pro menší soubory, které nemají mnoho stejných dvojic, je vhodnější zadání lineární tabulkou:

x	***y***
x_1	y_1
…	…
x_N	y_N

(příklad souboru, který je zadán lineární tabulkou)

	A	B	C	D
1				
2		x	y	
3		27	28	
4		31	21	
5		87	71	
6		93	36	
7		114	30	
8		124	43	
9		190	54	
10		193	54	
11		250	59	
12		254	25	
13		264	82	
14		272	22	
15		308	38	
16		324	22	
17		371	56	
18		372	63	
19		440	46	
20		442	24	
21		502	33	
22		503	40	
23		506	41	
24		522	28	
25		556	53	
26		620	38	
27		624	66	
28				

Momenty pak vypočteme jednodušeji:

$$m_{r,s} = \frac{1}{N}\sum_i x_i^r y_i^s$$

Centrální moment (*r* + *s*)-tého stupně je definován vztahem

$$n_{r,s} = \frac{1}{N}\sum_i\sum_j \left(x_i - m_{1,0}\right)^r \left(y_j - m_{0,1}\right)^s f_{i,j} = \sum_i\sum_j \left(x_i - m_{1,0}\right)^r \left(y_j - m_{0,1}\right)^s \varphi_{i,j}$$

Ze všech možných momentů se v podstatě používají jen prvé a druhé. Jejich význam už vlastně většinou známe:

$m_{1,0} = \bar{x}$ je **střední hodnota veličiny *x*** bez ohledu na chování veličiny *y*

$m_{0,1} = \bar{y}$ je **střední hodnota veličiny *y*** bez ohledu na chování veličiny *x*

$n_{2,0} = s_x^2$ je **rozptyl (variance) veličiny *x*** bez ohledu na rozptýlenost veličiny *y*

$n_{0,2} = s_y^2$ analogicky

Rozptýlenost obou veličin ve všech jejich vzájemných kombinacích postihuje smíšený moment druhého stupně

$n_{1,1} = \operatorname{cov} xy = \frac{1}{N}.\sum_i\sum_j f_{ij}\left(x_i - \bar{x}\right)\left(y_j - \bar{y}\right) = \frac{1}{N}.\sum_i\sum_j f_{ij}x_i y_j - \bar{x}.\bar{y}$ ... tzv. **kovariance**, jejíž normovaná bezrozměrná forma

$n_{1,1} = \frac{\operatorname{cov} xy}{s_x . s_y} = r$ je **koeficient (lineární) korelace**. Jeho význam a interpretaci poznáme v kapitole 9.

Přímý výpočet momentů lze pohodlně provést u momentů počátečních, takže je, obzvláště u ručního počítání, výhodné si odvodit vztahy:

$$n_{2,0} = m_{2,0} - m_{1,0}^2$$

$$n_{0,2} = m_{0,2} - m_{0,1}^2$$

$$n_{1,1} = m_{1,1} - m_{1,0}m_{0,1}$$

analogicky jako u momentů jednorozměrné náhodné veličiny.

Je-li soubor zadán lineární tabulkou pomocí dvojic (x_i, y_i), lze např. koeficient korelace vypočíst podle vzorce upraveného do tvaru:

$$r = \frac{N\sum x_i y_j - \sum x_i \sum y_j}{\sqrt{\left(N\sum x_i^2 - \left(\sum x_i\right)^2\right) . \left(N\sum y_j^2 - \left(\sum y_j\right)^2\right)}} .$$

Vícerozměrný statistický soubor velmi často charakterizujeme tzv. **kovarianční maticí**

$\begin{vmatrix} s_x^2 & \operatorname{cov} xy \\ \operatorname{cov} xy & s_y^2 \end{vmatrix}$, resp. její normovanou formou, **korelační maticí** $\begin{vmatrix} 1 & r \\ r & 1 \end{vmatrix}$.

Jejich důležitost však se projevuje hlavně v případě mnoharozměrných náhodných veličin.

Poznámka

Uvedené vzorce lze samozřejmě přímo použít k výpočtu definovaných veličin, ale je zřejmé, že programové vybavení současných počítačů skýtá daleko pohodlnější cestu, jak výsledky získat. Ideální je v tomto případě použití libovolného tabulkového kalkulátoru. Prostudujte si následující řešené příklady. Sledujte, jak se dá využít klasické tabelační činnosti excelu i pokročilejších technik při práci s tzv. maticovými operacemi.

Řešení příkladů, jejichž zadání jsme sledovali v textu:

Řešené úlohy

Příklad 8.1.1. Vypočtěte charakteristiky statistického souboru se dvěma argumenty. Zadání v Excelu:

	A	B	C	D	E	F	G	H	I
1									
2		**x \ y**	**20**	**30**	**40**	**50**	**60**	**70**	**80**
3		**250**	19	5					
4		**350**	23	116	11				
5		**450**	1	41	98	9			
6		**550**		4	32	65	7		
7		**650**		1	4	21	46	3	
8		**750**			1	2	11	13	1
9		**850**					1	3	2
10									
11									

Řešení: V excelu jsme vypočetli potřebné součty:

x \ y	20	30	40	50	60	70	80	fx	x*fx	x^2*fx	x*y*fx
250	19	5						24	6000	1500000	132500
350	23	116	11					150	52500	18375000	1533000
450	1	41	98	9				149	67050	30172500	2529000
550		4	32	65	7			108	59400	32670000	2788500
650		1	4	21	46	3		75	48750	31687500	2736500
750			1	2	11	13	1	28	21000	15750000	1342500
850					1	3	2	6	5100	4335000	365500
fy	43	167	146	97	65	19	3	540	259800	134490000	11427500
y*fy	860	5010	5840	4850	3900	1330	240	22030			
y^2*fy	17200	150300	233600	242500	234000	93100	19200	989900			
x*y*fy	265000	1894500	2756000	2747500	2571000	997500	196000	11427500			

Střední hodnoty:

$$\bar{x} = m_{1,0} = \frac{1}{N}.\sum_i x_i N_i = \frac{1}{540}.259800 \doteq 481,\bar{1}$$

$$\bar{y} = m_{0,1} = \frac{1}{N}.\sum_j y_j M_j = \frac{1}{540}.22030 \doteq 40,80$$

Rozptyly:

$${}^2_x = n_{2,0} = m_{2,0} - m_{1,0}^2 = \frac{1}{N}.\sum_i x_i{}^2 N_i - m_{1,0}^2 =$$

$$= \frac{1}{540}.134490000 - 481,1^2 \doteq 17587,65$$

$$s_y^2 = n_{0,2} = m_{0,2} - m_{0,1}^2 = \frac{1}{N}.\sum_j y_j{}^2 M_j - m_{0,1}^2 =$$

$$= \frac{1}{540}.989900 - 40,8^2 \doteq 168,81$$

Směrodatné odchylky:

$$s_x = \sqrt{17587,65} \doteq 132,62$$

$$s_y = \sqrt{168,81} \doteq 12,99$$

Kovariance:

$$\operatorname{cov} xy = n_{1,1} = \frac{1}{N}.\sum_i \sum_j f_{ij} x_i y_j - \bar{x}.\bar{y} =$$

$$= 11427500 - 481,1.40,8 \doteq 1534,49$$

Koeficient korelace:

$$r = \frac{\operatorname{cov} xy}{s_x s_y} = \frac{1534,49}{132,62.12,99} \doteq 0,891$$

Předchozí úlohu si můžete otevřít **vyřešenou** v Excelu.

Příklad 8.1.2. Vypočtěte číselné charakteristiky statistického souboru se dvěma argumenty, který je zadán lineární tabulkou:

x	27	31	87	93	114	124	190	193	250	254	264	272	308	324	371	372	440	442	502	503	506	522	556	620	624
y	28	21	71	36	30	43	54	54	59	25	82	22	38	22	56	63	46	24	33	40	41	28	53	38	66

Řešení: Vše potřebné opět vypočteme např. v Excelu:

	x	y	x²	y²	x*y
	27	28	729	784	756
	31	21	961	441	651
	87	71	7569	5041	6177
	93	36	8649	1296	3348
	114	30	12996	900	3420
	124	43	15376	1849	5332
	190	54	36100	2916	10260
	193	54	37249	2916	10422
	250	59	62500	3481	14750
	254	25	64516	625	6350
	264	82	69696	6724	21648
	272	22	73984	484	5984
	308	38	94864	1444	11704
	324	22	104976	484	7128
	371	56	137641	3136	20776
	372	63	138384	3969	23436
	440	46	193600	2116	20240
	442	24	195364	576	10608
	502	33	252004	1089	16566
	503	40	253009	1600	20120
	506	41	256036	1681	20746
	522	28	272484	784	14616
	556	53	309136	2809	29468
	620	38	384400	1444	23560
	624	66	389376	4356	41184
Σ	7989	1073	3371599	52945	349250

Střední hodnoty:

$$\bar{x} = m_{1,0} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{25}.7989 = 319{,}56$$

$$\bar{y} = m_{0,1} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{25}.1073 = 42{,}92$$

Rozptyly:

$$s_x^2 = n_{2,0} = m_{2,0} - m_{1,0}^2 = \frac{1}{N}.\sum_i x_i^{\;2} - m_{1,0}^2 =$$

$$= \frac{1}{25}.3371599 - 319,56^2 \doteq 32745,37$$

$$s_y^2 = n_{0,2} = m_{0,2} - m_{0,1}^2 = \frac{1}{N}.\sum_i y_i^{\;2} - m_{0,1}^2 =$$

$$= \frac{1}{25}.52945 - 42,92^2 \doteq 275,67$$

Směrodatné odchylky:

$$s_x = \sqrt{32745,37} \doteq 180,96$$

$$s_y = \sqrt{275,67} \doteq 16,60$$

Kovariance:

$$\operatorname{cov} xy = n_{1,1} = \frac{1}{N}.\sum_i \sum_j x_i y_j - \overline{x}.\overline{y} = (\text{v tomto případě}) =$$

$$= \frac{1}{N} \cdot \sum_i x_i y_i - \overline{x} \cdot \overline{y} = \frac{1}{25}.349250 - 319,56.42,92 \doteq 254,48$$

Koeficient korelace:

$$r = \frac{\operatorname{cov} xy}{s_x s_y} = \frac{254,48}{180,96.16,60} \doteq 0,085$$

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Poznámka

Při řešení předchozího příkladu jsme mohli použít i předdefinovaných funkcí v Excelu, jak bylo ukázáno v 6. kapitole, příkladu 6.2.1. nebo doplňkového nástroje Analýza dat *obdobným způsobem, jak bylo popsáno v 7. kapitole, příkladu 7.3.1.*

Poznámka

I když jsme se dosud věnovali zpracování statistického souboru, který jakoby byl realizací dvojrozměrné ***diskrétní*** *náhodné veličiny, je zřejmé, že práce se* ***spojitou*** *veličinou se nutně musí na tento případ převést. Realizace spojité veličiny se projeví vznikem číselné hodnoty zadané s určitou přesností nebo nějakým způsobem zaokrouhlené. Z praktických důvodů je také někdy vhodné hodnoty jednotlivých argumentů určitým způsobem setřídit, roztřídit do* *tříd a umožnit tak vlastně přechod k diskrétním veličinám reprezentovaným středy použitých tříd. A pak předešlé postupy jsou dokonale použitelné. Problém velikosti chyby, které se takovým zaokrouhlením dopouštíme, je ovšem nutno zohlednit. U jednorozměrného souboru jsou známé korekce, které s ohledem na šířku třídy umožní opravit vypočtené charakteristiky (Shepardovy korekce). U vícerozměrných šetření se takové korekce neprovádějí. Poznamenejme ještě, že v dnešní době, kdy zpracování statistických souborů stejně svěřujeme počítačům, není problém předběžné úpravy dat (např. tříděním a tedy zaokrouhlováním) tak podstatný, neboť počítačové postupy nejsou na množství nebo numerické "nevhodnosti" dat tak závislé a je možné pracovat přímo s prvotními daty.*

Úlohy k samostatnému řešení

8.1. U studentů 1.ročníku byly zaznamenány výsledky zkoušek z matematiky, fyziky a programování. Jsou uvedeny ve formě trojic číslic, z nichž první je známka z matematiky, druhá z fyziky a třetí z programování:

111	111	112	112	113	122	122	121	122	123
124	122	121	131	132	143	212	212	212	213
212	212	221	224	223	222	222	222	223	222
231	233	232	232	231	231	232	233	234	232
231	233	232	234	233	233	233	233	232	232
241	242	314	312	311	313	313	313	313	322
321	324	323	322	323	323	323	323	324	323
323	333	332	332	334	333	333	333	332	334
334	332	332	333	332	331	332	333	333	333
331	332	334	333	333	333	333	333	332	333
334	333	333	333	332	333	334	333	343	343
342	343	344	343	343	343	424	434	443	432
431	432	433	442	443	443	443	443	443	442
444	444	444	444	444					

a) Vytvořte statistický soubor s dvěma argumenty, z nichž *X* bude znamenat výsledek zkoušky z matematiky a *Y* výsledek zkoušky z fyziky a určete jeho charakteristiky.

b) Vytvořte statistický soubor s dvěma argumenty, z nichž *X* bude znamenat výsledek zkoušky z matematiky a *Y* výsledek zkoušky z programování a určete jeho charakteristiky.

8.2. U 130 zákrsků bylo zjištěno stáří stromu v letech (argument *X*) a sklizeň v jistém roce v kg (argument *Y*). Podle údajů v tabulce určete charakteristiky tohoto souboru.

X\Y	4	5	6	7	8	9	10	11
3	6	0	0	0	0	0	0	0
4	0	5	10	2	0	0	0	0
5	0	0	0	2	8	3	0	0
6	0	0	0	0	0	12	10	0
7	0	0	0	0	0	8	15	4
8	0	0	0	0	4	16	8	0
9	0	3	12	2	0	0	0	0

Výsledky úloh k samostatnému řešení

Výsledky:

8.1. *a*) $\bar{x} = 2{,}64; \bar{y} = 2{,}69; s_x^2 = 0{,}75; s_y^2 = 0{,}822; k_{xy} = 0{,}354; r_{xy} = 0{,}451;$

regresní přímky: $y = 0{,}472x + 1{,}445; x = 0{,}43y + 1{,}48$; $\Phi = 41^\circ 30'$;

$s_{\bar{x}}^2 = 0{,}1663; s_{\bar{y}}^2 = 0{,}1883; p_{yx} = 0{,}479; p_{xy} = 0{,}471$

b) $\bar{x} = 2{,}637; \bar{y} = 2{,}607; s_x^2 = 0{,}75; s_y^2 = 0{,}787; k_{xy} = 0{,}295; r_{xy} = 0{,}384$;

regresní přímky: $y = 0{,}393x + 1{,}571; x = 0{,}374y + 1{,}661$; $\Phi = 48^\circ$;

$s_{\bar{x}}^2 = 0{,}113; s_{\bar{y}}^2 = 0{,}121; p_{yx} = 0{,}392; p_{xy} = 0{,}388$

8.2. $\bar{x} = 6{,}53; \bar{y} = 8{,}15; s_x^2 = 3{,}1; s_y^2 = 3{,}59; k_{xy} = 1{,}11; r_{xy} = 0{,}34$;

regresní přímky: $y = 0{,}37x + 5{,}74; x = 0{,}31y + 4{,}02$; $\Phi = 53^\circ$;

$s_{\bar{x}}^2 = 0{,}75; s_{\bar{y}}^2 = 3{,}24; p_{yx} = 0{,}95; p_{xy} = 0{,}5$

9. REGRESNÍ A KORELAČNÍ ANALÝZA

Průvodce studiem

V předchozí kapitole jsme uvedli způsob, jak popsat lineární závislost mezi dvěma argumenty a její míru. Užitím korelačních poměrů je možné zjistit, zda má smysl hledat jiný typ závislosti mezi proměnnými než lineární.

Předpokládané znalosti

Pojmy z předchozích kapitol.

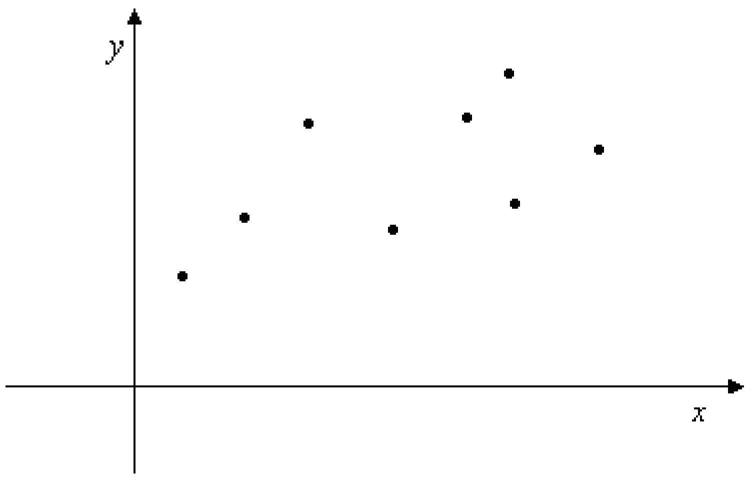
Cíle

Cílem této kapitoly je vysvětlit pojmy regrese, korelace, regresní funkce, metoda nejmenších čtverců odchylek, index korelace.

Výklad

9.1. Lineární regrese

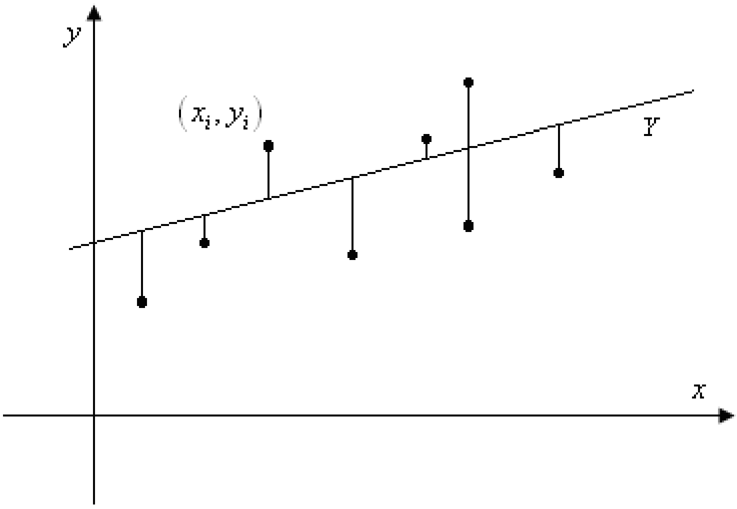
Grafické zobrazení dvojrozměrné náhodné veličiny, statistický soubor s dvěma statistickými znaky (x_i,y_i); $i = 1,2,...,n$ (korelační pole):



Hledejme vyjádření této "**statistické**" závislosti "nejlepším" **funkčním** předpisem. A pro začátek předpokládejme tento předpis lineární:

$$Y = a + bx$$

Jako kritérium pro "nejlepší" funkční předpis vezměme z určitých důvodů (známých už např. Gaussovi v počtu pravděpodobnosti i např. proto, že se takový přístup úspěšně uplatňuje i v jiných situacích – viz. ukázka – pouze na webu) minimalizaci sumy kvadrátů odchylek empirických hodnot y od teoretických hodnot získaných pomocí předpisu y_t:



$$S(a,b) = \sum_{i=1}^{n} (Y_i - y_i)^2 = \sum_{i=1}^{n} (a + bx_i - y_i)^2 = \min$$

Hodnota veličiny S závisí na volitelných hodnotách a a b a je to tedy funkce dvou proměnných. Její extrém se najde nulováním parciálních derivací podle těchto proměnných.

$$\frac{\partial S}{\partial a} = 2.\sum_{i=1}^{n} (a + bx_i - y_i).1 = 0$$

$$\frac{\partial S}{\partial b} = 2.\sum_{i=1}^{n} (a + bx_i - y_i).x_i = 0$$

Po úpravě dojdeme k soustavě lineárních rovnic pro určení a a b. (V dalším textu budeme někdy zjednodušovat zápis sumační symboliky.)

$$n.a + b.\sum_i x_i = \sum_i y_i$$

$$a.\sum_i x_i + b.\sum_i x_i^2 = \sum_i x_i y_i$$

Tuto soustavu můžeme vyřešit mnoha způsoby. Například pomocí determinantu matice soustavy, který lze upravit na vyjádření pomocí rozptylu:

$$D = n.\sum_i x_i{}^2 - \left(\sum_i x_i\right)^2 = n^2.s_x^2,$$

takže koeficienty rovnice přímky nakonec jsou:

$$a = \frac{n \cdot \sum_i y_i \cdot \sum_i x_i{}^2 - \sum_i x_1 \cdot \sum_i x_i y_i}{n^2 \cdot s_x{}^2}$$

$$b = \frac{n.\sum_i x_i y_i - \sum_i x_i.\sum_i y_i}{n^2.s_x^2}$$

Po poněkud pracnějších úpravách (s využitím vyjádření centrálních momentů pomocí momentů počátečních):

$$Y = \frac{\sum_i y_i \cdot \sum_i x_i^2 - \sum_i x_i \cdot \sum_i x_i y_i}{n^2.s_x^2} + \frac{n.\sum_i x_i y_i - \sum_i x_i \cdot \sum_i y_i}{n^2.s_x^2}.x$$

$$Y = \frac{1}{s_x^2}.\left(\frac{\sum_i y_i}{n}.\frac{\sum_i x_i^2}{n} - \frac{\sum_i x_i}{n}.\frac{\sum_i x_i \cdot y_i}{n} + \frac{\sum_i x_i \cdot y_i}{n}.x - x.\frac{\sum_i x_i}{n}.\frac{\sum_i y_i}{n}\right)$$

$$Y = \frac{1}{s_x^2}.\left(\bar{y}.\frac{\sum_i x_i^2}{n} - \bar{y}.\left(\frac{\sum_i x_i}{n}\right)^2 + \bar{y}.\left(\frac{\sum_i x_i}{n}\right)^2 - \bar{x}.\frac{\sum_i x_i \cdot y_i}{n} + x.\frac{\sum_i x_i \cdot y_i}{n} - x.\bar{x}.\bar{y}\right)$$

$$Y = \frac{1}{s_x^2}.\left(\bar{y}.s_x^2 + \frac{\sum_i x_i \cdot y_i}{n}\ \ x - \bar{x}\ \ - \bar{x}.\bar{y}.\ \ x - \bar{x}\right)$$

$$Y = \bar{y} + \frac{1}{s_x^2}.\left(\frac{\sum_i x_i \cdot y_i}{n} - \bar{x}.\bar{y}\right).\ \ x - \bar{x}$$

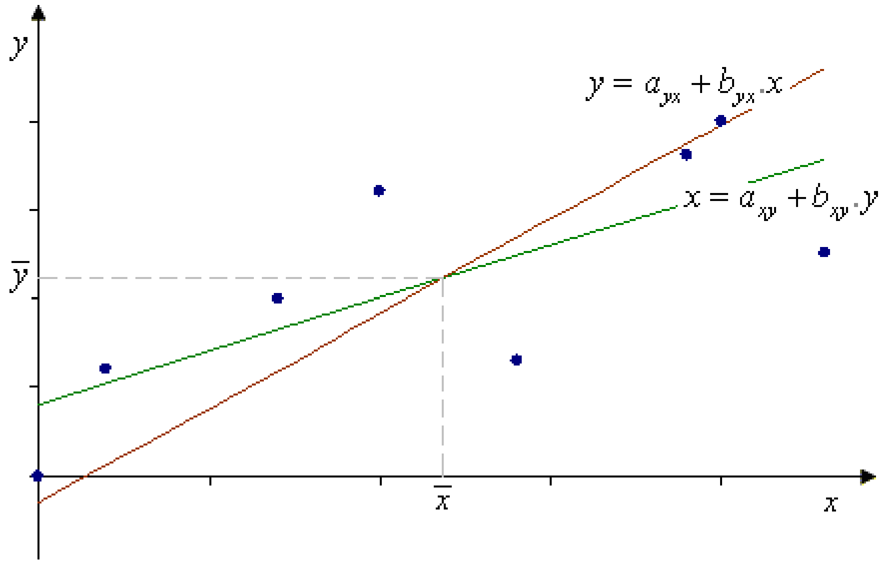
dostáváme jinou podobu **rovnice regresní přímky**, z níž vyplývá, že tato přímka prochází tzv. centrálním bodem $\left[\bar{x}, \bar{y}\right]$ ($\bar{x}$, $\bar{y}$ jsou střední hodnoty proměnných x, y) a že směrnici přímky, tzv. koeficient regrese, ovlivňuje jak kovariance, tak rozptyl té proměnné, která byla prohlášena za nezávislou:

$$y-\bar{y}=\frac{\operatorname{cov} xy}{s_x^2}.\ x-\bar{x}$$

Tuto volbu můžeme pochopitelně změnit a tak se dojde analogickou cestou k jiné regresní přímce:

$$x-\bar{x}=\frac{\operatorname{cov} xy}{s_y^2}.\ y-\bar{y}$$

Vykreslíme-li obě takto získané přímky do jedné souřadnicové soustavy, dostaneme tzv. regresní nůžky:



Směrnice obou regresních přímek $b_{yx}=\frac{\operatorname{cov} xy}{s_x^2}$ a $b_{xy}=\frac{\operatorname{cov} xy}{s_y^2}$ nazýváme **regresní koeficienty** při závislosti y na x, resp. x na y a mají velmi důležitou praktickou interpretaci: udávají **přírůstek závisle proměnné při jednotkové změně nezávisle proměnné**. (Dokažte!) Zároveň umožňují vypočíst koeficient lineární korelace, který jsme výše definovali jako normovaný smíšený moment druhého stupně, vypočíst jiným způsobem:

$$b_{yx}.b_{xy} = \frac{(\operatorname{cov} xy)^2}{s_x^2.s_y^2} = r^2$$

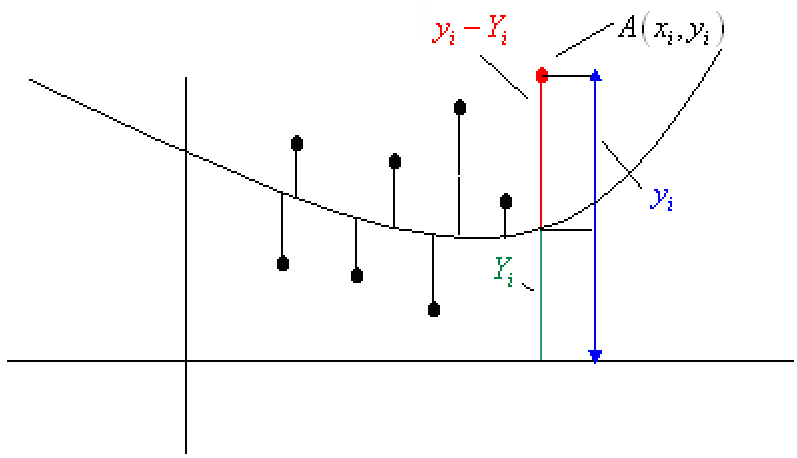
Znaménko přidělíme podle znaménka kteréhokoliv regresního koeficientu, např.:

$$r = \operatorname{sign}(b_{yx}).\sqrt{b_{yx}.b_{xy}}$$

Dá se dokázat, že tento koeficient nabývá hodnoty z intervalu $\langle -1,1 \rangle$ a měří vhodnost lineární funkce vyjádřit statistickou závislost mezi veličinami x a y. Čím je hodnota koeficientu blíže krajním hodnotám, tím je náhrada těsnější. V případě, že tento koeficient nabývá hodnoty 1 nebo -1, leží všechny body na regresní přímce a závislost veličin x a y je přesně lineární.

Stanovit stupnici oceňující závislost (závislost "slabá", "střední", "silná") není úkol pro matematika, ale pro profesního odborníka. Podobné stupnice bývají součástí oborových norem.

Lineární průběh nemusí vždy vystihovat vzájemné chování obou složek dvojrozměrné náhodné veličiny. Nic ale nestojí v cestě přirozenému zobecnění předešlých úvah a postupů.



Uvažujme jako výše korelační pole (x_i, y_i); $i = 1,2,...,n$ a funkci (kterou volíme pouze jejím charakterem, ale nikoliv jejími parametry, které určují detailně průběh funkce) $Y = f(x, a_0, a_1, \ldots, a_k)$, která by měla vyjádřit vztah mezi složkami x a y. A hledejme množinu koeficientů a_i tak, aby byl splněn požadavek MNČ (metody nejmenších čtverců):

$$S\left(x,a_0,a_1,\ldots,a_k\right)=\sum_{i=1}^{n}\left[f\left(x,a_0,a_1,\ldots,a_k\right)-y_i\right]^2=\min$$

Řešením soustavy rovnic:

$$\frac{\partial S\left(x,a_0,a_1,\ldots,a_k\right)}{\partial a_j}=0;\ \ j=0,...,k\,,$$

vzniklé nulováním parciálních derivací funkce S podle jednotlivých hledaných koeficientů, dostaneme hledanou regresní funkci. Mohou však nastat problémy algebraického charakteru. Vzniklá soustava rovnic může být velmi nesnadno řešitelná (zvlášť bez použití výpočetní techniky). Proto se zpravidla hledají vhodné regresní funkce pouze mezi tzv. adičními funkcemi:

$$f\left(x,a_0,a_1,\ldots,a_k\right)=a_0+a_1.f_1\left(x\right)+\ldots+a_k.f_k\left(x\right)$$

Ty totiž vedou k řešení soustavy lineárních rovnic, jak lze snadno ukázat. Na případy adičních funkcí se často převádějí i funkce multiplikativní, jako je např. funkce mocninná či exponenciální. Linearizace logaritmováním funkčního předpisu však obecně dává pouze suboptimální řešení z hlediska MNČ.

Postup ukážeme na regresní funkci

$$Y = a.\mathrm{e}^{bx}$$

Tuto funkci použijeme za předpokladu, že rychlost růstu závisle proměnné je přímo úměrná její velikosti.

Při určování konstant a, b zlogaritmujeme funkci:

$$\ln Y = \ln a + bx$$

Jestliže nyní položíme $Z = \ln Y$, $a_1 = \ln a$, je funkce

$$Z = a_1 + bx$$

lineární v parametrech a můžeme použít již známého postupu. Hledáme tedy minimum funkce

$$\sum_i \left(a_1+bx_i-z_i\right)^2.$$

Po sestavení soustavy rovnic se můžeme vrátit k původním proměnným. Soustava bude mít tedy tvar:

$$N \cdot \ln a + b \cdot \sum_i x_i = \sum_i \ln y_i$$
$$\ln a \cdot \sum_i x_i + b \cdot \sum_i x_i^2 = \sum_i x_i \cdot \ln y_i$$

Podobně postupujeme např. pro funkci $Y = a.x^b$ (kde b není přirozené číslo) nebo $Y = \dfrac{1}{a + b \cdot \Phi(x)}$ (v tomto případě lze použít transformace $Z = \dfrac{1}{Y}$).

Poznámka

Hledisko numerické náročnosti regresní analýzy se stává v současné době druhořadé, neboť standardní počítačové programy nabízejí automatizované řešení této úlohy.

Podstatnější problém nastává při měření vhodnosti regresní funkce. Koeficient lineární korelace tu ztrácí svůj význam a je třeba najít jinou míru těsnosti uvažovaného vztahu a daného korelačního pole.

Zaveďme tato označení pro speciálním způsobem definované rozptyly:

$$s_y^2 = \frac{1}{n}.\sum_i \left(y_i - \bar{y}\right)^2$$

$$s_Y^2 = \frac{1}{n}.\sum_i \left(Y_i - \bar{y}\right)^2$$

$$s_{y.x}^2 = \frac{1}{n} \cdot \sum_i \left(y_i - Y_i\right)^2,$$

když Y_i je funkční hodnota regresní funkce příslušná i-té x-ové složce.
Všimněme si, jaký mezi nimi existuje vztah:

$$s_y^2 = \frac{1}{n}.\sum \left(y_i - \bar{y}\right)^2 = \frac{1}{n}.\sum \left(\left(y_i - Y_i\right) + \left(Y_i - \bar{y}\right)\right)^2$$

$$= \frac{1}{n}.\sum \left(\left(y_i - Y_i\right)^2 + \left(Y_i - \bar{y}\right)^2 + 2.\left(y_i - Y_i\right).\left(Y_i - \bar{y}\right)\right) =$$

$$= s_{y.x}^2 + s_Y^2 + \frac{2}{n}.\sum \left(y_i - Y_i\right).\left(Y_i - \bar{y}\right)$$

Dá se dokázat (ukázka pouze na webu), že poslední výraz na pravé straně je roven nule.

Pak $s_y^2 = s_{yx}^2 + s_Y^2$ a podíl $\frac{s_Y^2}{s_y^2} = 1 - \frac{s_{yx}^2}{s_y^2} \in \langle 0;1 \rangle$ bývá používán jako míra těsnosti, vhodnosti regresní funkce (**koeficient determinace**). Udává vlastně, jaká část disperze znaku *y* je způsobena závislostí na *x*. Doplněk koeficientu determinace do jedné znamená podíl náhodné složky na disperzi. Odmocnina $I_{yx} = \frac{s_Y}{s_y} = \sqrt{1 - \frac{s_{yx}^2}{s_y^2}}$ (**index korelace**) má analogickou interpretaci jako koeficient korelace (pro lineární regresní vztah jde o zcela totožný výsledek).

Poznámka

K posouzení míry vhodnosti regresní funkce může sloužit také pouze hodnota

$s_{y.x}^2 = \frac{1}{n} \cdot \sum_i \left(y_i - Y_i\right)^2$ - **reziduální (zbytkový) součet čtverců (rozptyl)**. *Nejvhodnější regresní funkcí je pak samozřejmě ta funkce, která má reziduální součet čtverců nejnižší.*

Řešené úlohy

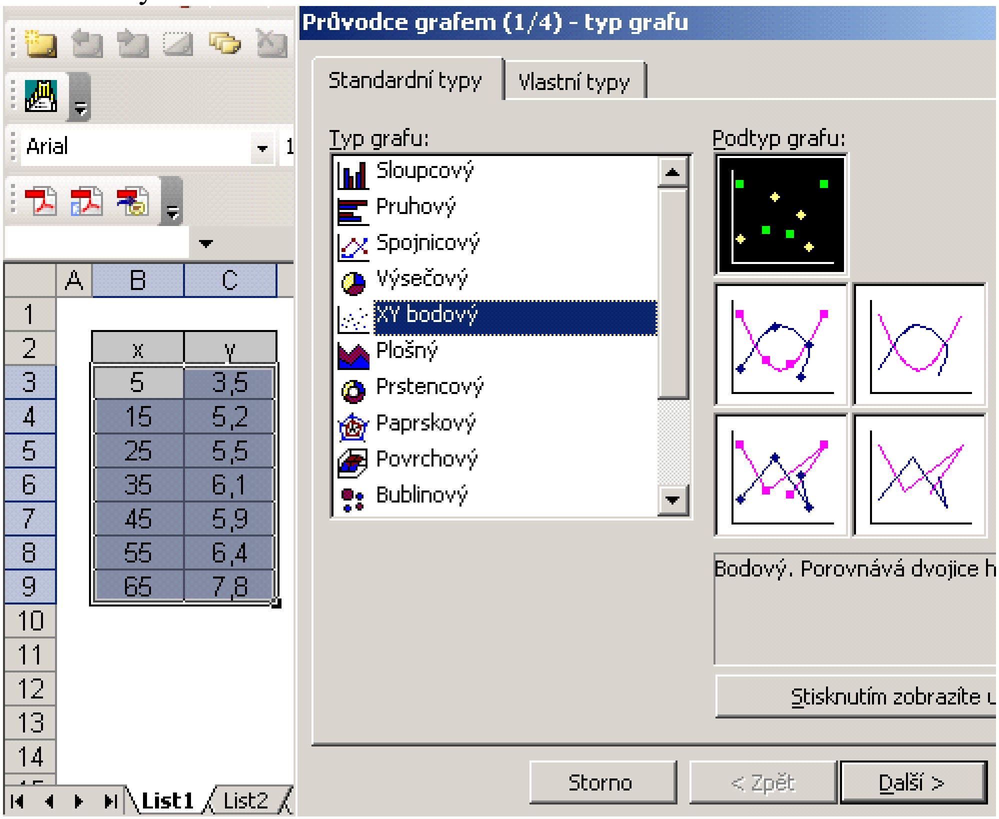
Příklad 9.1.1. Vyrovnejte data v tabulce regresní přímkou

x	5	15	25	35	45	55	65
y	3,5	5,2	5,5	6,1	5,9	6,4	7,8

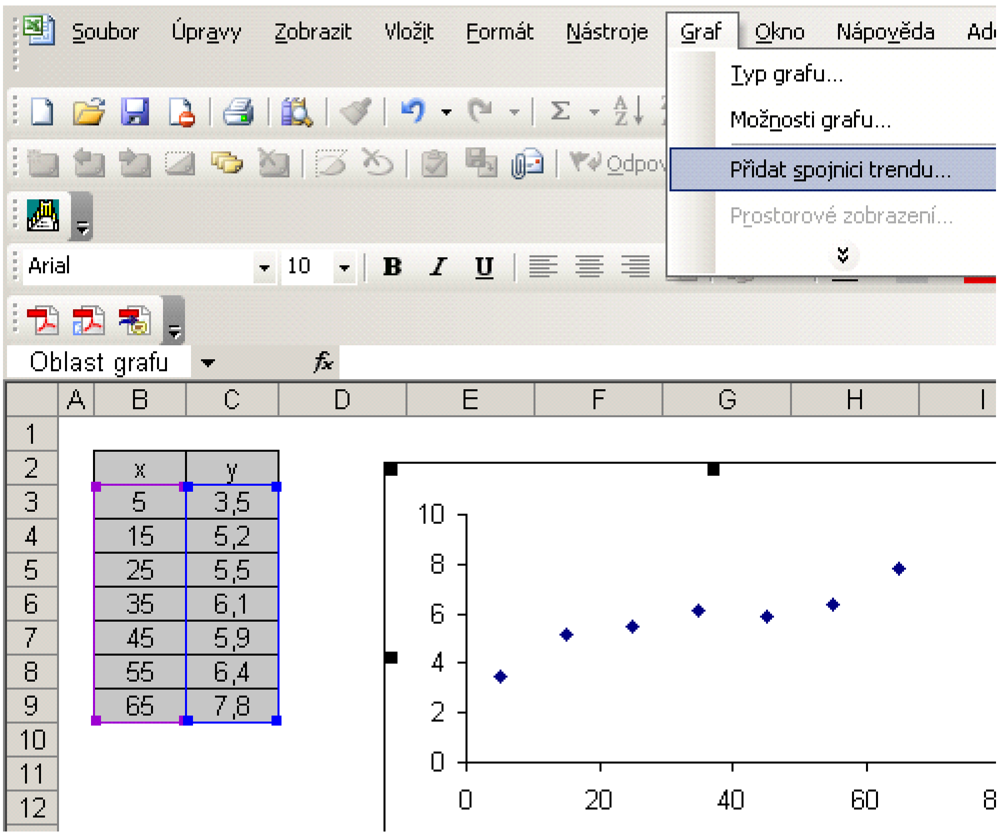
Řešení: Ukážeme, jak by se tato úloha řešila v Excelu:

Nejdříve označíme data a klikneme na Vložit Graf..., přičemž vybereme typ grafu

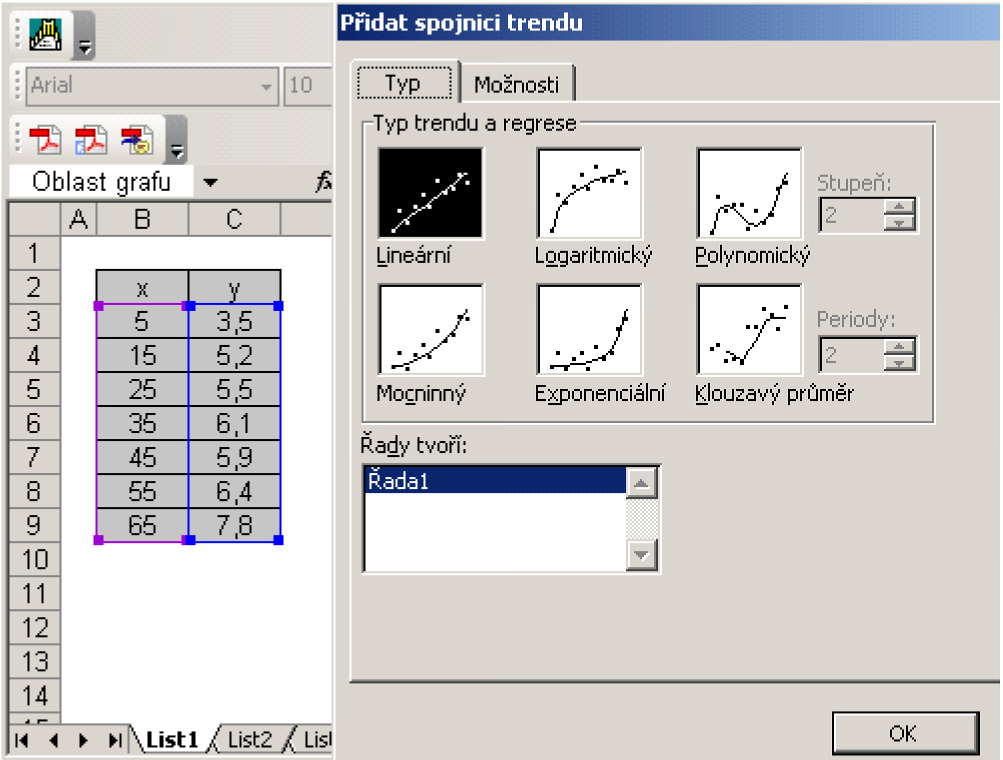
XY bodový:



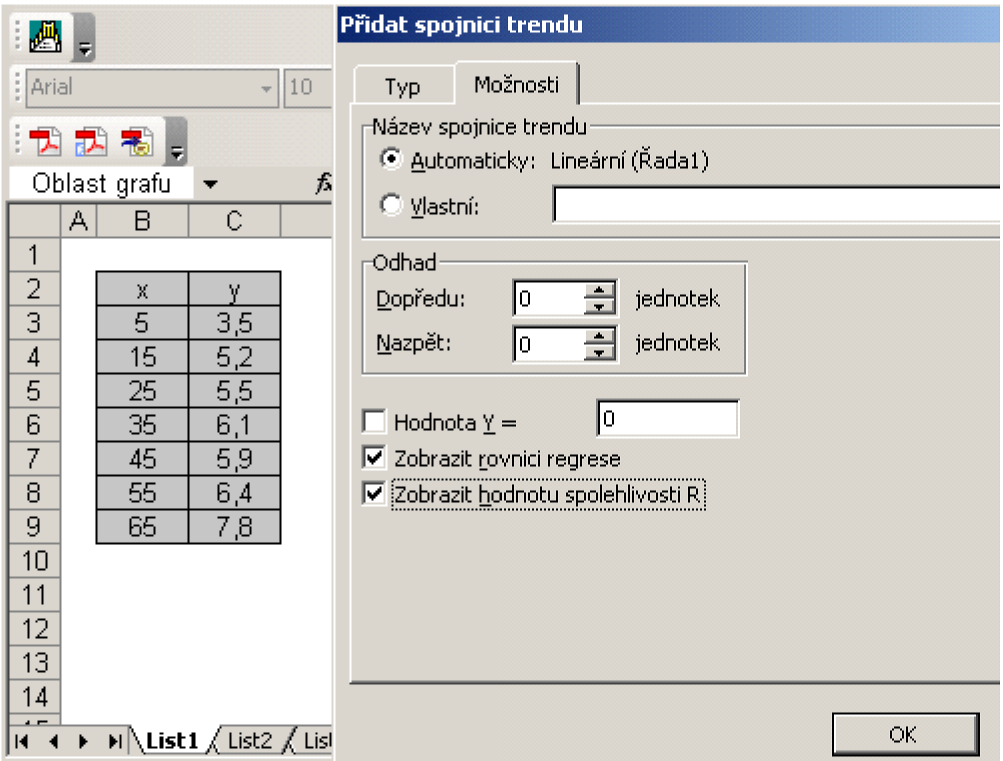
Máme-li aktivní okno grafu, v nabídce Excelu přibude položka Graf, vybereme možnost Přidat spojnici trendu...:



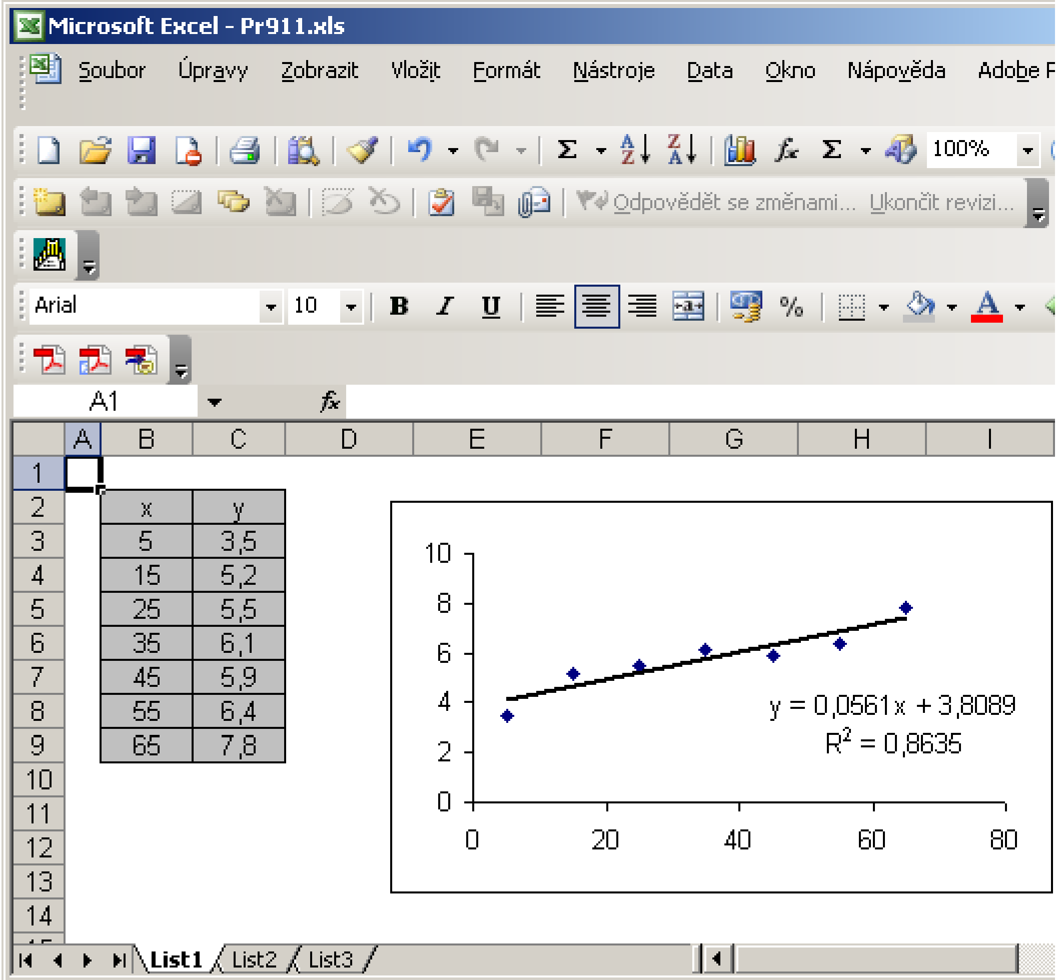
Chceme-li daty proložit přímku, vybereme Typ trendu - lineární:



Pro zobrazení rovnice regrese a hodnoty spolehlivosti R (druhá mocnina indexu korelace) klikneme na kartu Možnosti a zaškrtneme příslušné položky:



Konečná podoba řešení:



Z grafu vidíme, že rovnice regrese je: $y = 0{,}0561.x + 3{,}8089$, index korelace:

$$I_{yx} = \sqrt{0{,}8635} = 0{,}9292$$

V tomto případě existuje i další možnost, jak vypočíst koeficienty *a*, *b* v rovnici regrese a index korelace. Rovnici regrese vypočteme pomocí v Excelu předdefinované funkce LINREGRESE, kterou najdeme v kategorii statistické. Nutno mít na paměti, že výsledkem budou dvě hodnoty, proto před vyvoláním této funkce označíme dvě buňky vedle sebe a při použití stiskneme současně klávesy CTRL+SHIFT+ENTER (matice na výstupu). V našem příkladě by se tato funkce zadávala takto: LINREGRESE(C3:C9;B3:B9;1).

Index korelace je v tomto případě shodný s koeficientem korelace (viz. kapitola 8), tudíž použijeme předdefinovanou funkci: CORREL(B3:B9;C3:C9)

Předchozí úlohu si můžete otevřít **vyřešenou** v Excelu.

Poznámka

Na druhém listě řešení předchozího příkladu v Excelu je provedena regresní analýzu pomocí doplňkového nástroje Analýza dat *(použití popsáno v 7. kapitole, příkladu 7.3.1.), analytický nástroj* Regrese.

Poznámka

Jak je patrné z třetího obrázku v řešení předchozího příkladu, obdobně bychom postupovali v případě, že bychom potřebovali daty proložit např. logaritmickou, exponenciální, mocninnou funkci, případně polynom 2.-6. stupně.

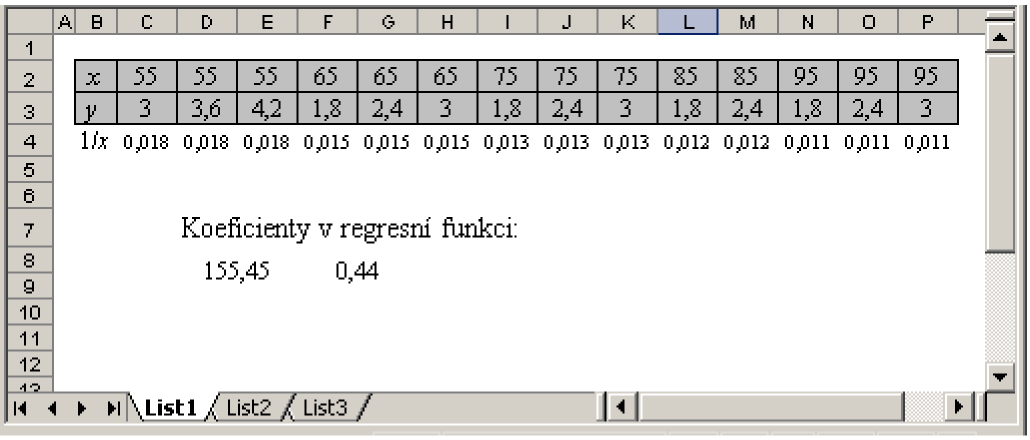
Řešené úlohy

Příklad 9.1.2. Charakterizujte závislost proměnné *y* na *x* regresní funkcí ve tvaru hyperboly

$$y = a + \frac{b}{x}$$

x	55	55	55	65	65	65	75	75	75	85	85	95	95	95
y	3	3,6	4,2	1,8	2,4	3	1,8	2,4	3	1,8	2,4	1,8	2,4	3

Řešení: Úlohu vyřešíme opět v Excelu, použijeme obdobně jako v předchozím příkladě předdefinovanou funkci LINREGRESE, která počítá koeficienty v lineární regresní funkci *y* = *a.x* + *b*. Pouze místo proměnné *x* do této rovnice dosadíme proměnnou $\frac{1}{x}$:



x	55	55	55	65	65	65	75	75	75	85	85	95	95	95
y	3	3,6	4,2	1,8	2,4	3	1,8	2,4	3	1,8	2,4	1,8	2,4	3
1/x	0,018	0,018	0,018	0,015	0,015	0,015	0,013	0,013	0,013	0,012	0,012	0,011	0,011	0,011

Koeficienty v regresní funkci:

155,45 0,44

Tato funkce je v tomto příkladě konkrétně zadána LINREGRESE(C3:P3;C4:P4;1)

Řešením je tedy regresní křivka ve tvaru hyperboly: $y = 0,44 + \frac{155,45}{x}$

Podobným způsobem vypočteme index korelace: CORREL(C3:P3;C4:P4). Index korelace je tedy roven: $I_{yx} = 0{,}608$.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Poznámka

Podobně bychom mohli samozřejmě hledat koeficienty v dalších regresních funkcích ve tvaru ve tvaru $y = a.f(x) + b$ *(např.* $y = a.x^3 + b$*).*

V rámci cvičení se věnujte následujícím úlohám:

- nalezení regresní přímky při standardním zadání souboru bodů (x_i, y_i) (postup při řešení v Excelu)
- nalezení regresní přímky při zadání dvojrozměrného souboru četnostní tabulkou (dokončete řešení příkladu z minulé kapitoly)
- nalezení nelineární regresní funkce podle nabídky kalkulátoru Excel
- nalezení nelineární regresní funkce podle MNČ bez předešlé linearizace (užitím numerického řešení, které nabízí *řešitel* Excelu (exponenciála, mocninná funkce)
- hledání zadání úloh z odborné profese čtenáře, které by vedly na regresní analýzu

Úlohy k samostatnému řešení

9.1. Charakterizujte závislost proměnné y na x regresní funkcí ve tvaru $Y = a + bx$

x	5	15	25	35	45	55	65
y	3,5	5,2	5,5	6,1	5,9	6,4	7,8

9.2. Charakterizujte závislost proměnné y na x regresní funkcí ve tvaru:

a) $Y = a + \frac{b}{x}$

b) $Y = ax^2 + bx + c$

Určete indexy korelace

x	1	1	3	4	6
y	0	1	4	5	5

9.3. Při seskoku parašutisty byla měřena závislost mezi rychlostí v [m/s] a tlakem p [0,1mPa] na povrchu padáku. Výsledky vyrovnejte parabolou $p = a + bv^2$. Vypočtěte index korelace.

v	2,4	3,5	5	6,89	10
p	0,0141	0,0281	0,0562	0,1125	0,225

9.4. Charakterizujte těsnost zvolené závislosti ve tvaru $Y = a + b.\log x$ mezi proměnnými x a y. Vypočtěte index korelace.

x	1	1	3	3	5	6	7	7
y	70	104	162	210	200	250	240	260

9.5. Při zjišťování závislosti veličin x a y byly naměřeny hodnoty uvedené v tabulce. Určete vhodnou regresní funkci.

x	55	55	55	65	65	65	75	75	75	85	85	95	95	95
y	3	3,6	4,2	1,8	2,4	3	1,8	2,4	3	1,8	2,4	1,8	2,4	3

9.6. Zjišťovalo se, zda u souboru chlapců je závislost v počtu provedených shybů a kliků. Výsledky jsou zaznamenány v tabulce:

chlapec	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
počet shybů	1	3	2	0	5	6	1	4	3	5	6	2	1	1	8
počet kliků	10	15	15	0	40	25	7	31	30	35	41	10	14	9	64

a) Určete, zda je mezi počtem shybů a počtem kliků silná lineární závislost, určete její míru.

b) Najděte nejvhodnější regresní funkci závislosti mezi počtem shybů a kliků.

Výsledky úloh k samostatnému řešení

9.1. $y = 0{,}056 + 3{,}809$

9.2. ***a)*** $Y = 6{,}06 - \dfrac{5{,}565}{x}$; $I = 0{,}985$; ***b)*** $Y = -2{,}15 + 2{,}942x - 0{,}2913x^2$; $I = 0{,}99$

9.3. $p = 0{,}00144 + 0{,}0022506v^2$; $I = 0{,}9996$

9.4. $Y = 88{,}32 + 191{,}54.\log x$; $I = 0{,}96$

9.5. $Y = 0{,}44 + \dfrac{155{,}43}{x}$

9.6. Lineární funkce: $y = 6{,}6939x + 1{,}6463$; $I_{yx} = 0{,}927577$

Kvadratická funkce: $y = 0{,}243x^2 + 4{,}8667x + 3{,}7354$; $I_{yx} = 0{,}93043$

10. ČASOVÉ ŘADY

Průvodce studiem

Využijeme znalostí z předchozích kapitol, především z 9. kapitoly, která pojednávala o regresní analýze, a rozšíříme je.

Předpokládané znalosti

Pojmy z předchozích kapitol.

Cíle

Cílem této kapitoly je seznámit s typy časových řad, jejich složkami a možnostmi analýzy časových řad.

Výklad

10.1. Časové řady - základní pojmy

Důležitými statistickými daty, pomocí nichž můžeme zkoumat dynamiku jevů v čase, jsou tzv. časové řady. Mají základní význam pro analýzu příčin, které na tyto jevy působily a ovlivňovaly jejich chování v minulosti, tak pro předvídání jejich budoucího vývoje.

Definice 10.1.1.

Časová řada (dynamická řada, vývojová řada)

je posloupnost pozorování kvantitativní charakteristiky uspořádaná v čase od minulosti do přítomnosti.

Podle Segera (viz seznam literatury) lze uvažovat o třech typech řad

1. časová řada intervalových ukazatelů
2. časová řada okamžikových ukazatelů
3. časová řada odvozených charakteristik

Pro ukazatele 1. typu platí, že jejich velikost přímo úměrně závisí na zvolené délce intervalu. (Uveďte příklady.) V těchto případech se často musí data převést na srovnatelné hodnoty (např. přepočet na stejně dlouhé úseky (čtvtletí nemají stejný počet dní apod.)).

U řad 2. typu se ukazatel vztahuje k přesně definovanému okamžiku. Hodnota ukazatele tedy nezávisí na délce intervalu, za který je sledován. Práce s těmito řadami je složitější. Na rozdíl od předešlého typu nemá reálný smysl např sumace hodnot řady, přistupuje se tedy k různým druhům průměrování.

Často je používán tzv. **chronologický průměr**:

$$\bar{x} = \frac{\frac{1}{2}x_1 + x_2 + \ldots + x_{n-1} + \frac{1}{2}x_n}{n-1}$$

Tímto jediným číslem pak charakterizujeme úroveň ukazatele za celé období. Je ale zřejmé, že tím dochází ke značnému zjednodušování reality. Oblíbenější jsou proto různé druhy klouzavých ukazatelů, které jsou schopny částečně eliminovat vliv náhodných vlivů na sledovaný ukazatel a tím časovou řadu "vyhladit". Používají se jak **klouzavé mediány**, tak **klouzavé průměry**. Vždy se postupuje tak, že udaj časové řady nahradíme zvoleným ukazatelem z okolních časově předcházejících a následujících údajů.

Poznámka

Zpracování časových řad užitím MS Excelu je zcela triviální. Způsob tvorby klouzavých ukazatelů je filozofii tabelárních výpočtů zcela přizpůsoben. A pokud jde o klouzavé průměry, disponuje excel přímo vestavěnou možností tyto ukazatele získat (analogický postup jako u regresní analýzy - viz ukázka – *pouze na webu).*

Řady 3. typu jsou odvozovány na základě absolutních údajů okamžikových nebo intervalových. Příkladem mohou být **časové řady součtové** nebo **časové řady poměrných čísel**

Při klasické analýze časových řad se vychází z předpokladu, že každá časová řada může obsahovat čtyři složky:

- trend,
- sezónní složku,
- cyklickou cložku,
- náhodnou složku.

Definice 10.1.2.

Trend

je obecná tendence vývoje zkoumaného jevu za dlouhé období. Je výsledkem dlouhodobých a stálých procesů. Trend může být rostoucí, klesající nebo může existovat řada bez trendu.

Sezónní složka

je pravidelně se opakující odchylka od trendové složky. Perioda této složky je menší než celková velikost sledovaného období.

Cyklická složka

udává kolísání okolo trendu v důsledku dlouhodobého cyklického vývoje (požíváno spíše v makroekonomických úvahách).

Náhodná (stochastická) složka

se nedá popsat žádnou funkcí času. "Zbývá" po vyloučení trendu, sezónní a cyklické složky.

Než přejdeme k analýze trendu a sezónnosti (dlouhodobou cykličnost ponecháme stranou našich úvah), uveďme několik jednoduchých ukazatelů, které se používají jako **míry dynamiky**:

absolutní přírůstek

$$\Delta y_t = y_t - y_{t-1}, \; t = 2, 3, \cdots, n$$

průměrný absolutní přírůstek

$$\overline{\Delta} = \frac{\sum \Delta y_t}{n-1} = \frac{y_2 - y_1 \;+\; y_3 - y_2 \;+ \cdots +\; y_n - y_{n-1}}{n-1} = \frac{y_n - y_1}{n-1}$$

relativní přírůstek

$$\delta_t = \frac{\Delta y_t}{y_{t-1}} = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1$$

průměrný koeficient růstu

$$\overline{k} = \sqrt[n-1]{k_1 k_2 \ldots k_n} = \sqrt[n-1]{\frac{y_2}{y_1}\frac{y_3}{y_3}\frac{y_4}{y_3}\cdots\frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Řešené úlohy

Příklad 10.1.1. Určete elementární charakteristiky růstu časové řady sledující výrobu plynu v letech 1980 - 1985.

rok	1980	1981	1982	1983	1984	1985
výroba (m^3)	1286	1363	1393	1495	1571	1610

Řešení:

rok	výroba (m^3) y_t	absolutní přírůstky	koeficienty růstu
1980	1286		
1981	1363	77	1,060
1982	1393	30	1,022
1983	1495	102	1,073
1984	1571	76	1,051
1985	1610	39	1,025

průměrný absolutní přírůstek:

$$\overline{\Delta} = \frac{\sum \Delta y_t}{n-1} = \frac{y_2 - y_1 + y_3 - y_2 + \cdots + y_n - y_{n-1}}{n-1} = \frac{y_n - y_1}{n-1} = 64{,}8$$

průměrný koeficient růstu:

$$\overline{k} = \sqrt[n-1]{k_1 k_2 \ldots k_n} = \sqrt[n-1]{\frac{y_2}{y_1}\frac{y_3}{y_3}\frac{y_4}{y_3}\ldots\frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}} = 1{,}046$$

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

10.2. Analýza trendu a sezónní složky

Nejčastěji se při analýze časové řady předpokládá **aditivní model** popisu chování řady. Předpokládá se, že jednotlivé složky vývoje se sčítají, takže platí:

$y_y = T_t + S_t + C_t + \varepsilon_t$, kde na pravé straně po řadě vystupují složky trendová, sezónní, cyklická a náhodná. Různé modifikace modelů vzniknou, když některou složku z úvah vypustíme. My tak učiníme pro složku cyklickou a o náhodné složce řekněme jen tolik, že o ní lze zpravidla předpokládat, že jejich střední hodnoty jsou nulové a že jsou korelačně nezávislé (náhodná porucha, jak se také dá náhodná složka interpretovat, nezávisí na poruše v minulém okamžiku ani neovlivňuje vznik a velikost poruchy v okamžiku následujícím).

Analýza složky kterékhokoliv typu se provádí v podstatě klasickou regresní analýzou. Podstatný rozdíl je jen v tom, že nezávisle proměnná, je v tomto případě proměnná časová a můžeme ji vcelku libovolně vyjádřit v jakýchkoliv časových jednotkách s libovolným počátkem.

Analýza trendové složky je zřejmě nejdůležitější částí analýzy časových řad. V průběhu let se potvrdilo, že při výběru trendových funkcí většinou vystačíme s úzkou nabídkou funkcí. Nejčastěji používané jsou

lineární trend	$y_t = a_0 + a_1 t$	Parametr a_1 představuje přírůstek hodnoty y připadající na jednotkovou změnu časové proměnné.
polynomický trend	$y_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k$	Umožňuje najít trendovou funkcí, která má extrém.
exponenciální trend	$y_t = a_0 a_1^t$	Parametr a_1 představuje průměrný přírůstek hodnot y_t. (Ty se chovají jako členy geometrické posloupnosti. Doložte vzpomínkami na tuto kapitolu středoškolské matematiky.)
modifikovaný exponenciální trend	$y_t = k + a_0 a_1^t$	Funkce má vodorovnou asymptotu a dá se pomocí ní snáze modelovat vývoj jevů, které vycházejí z omezených zdrojů růstu a u kterých existuje určitá mez nasycení, daná např. zájmem nebo potřebou určitého výrobku. (Předveďte si průběh funkcí tohoto typu pro různé hodnoty parametrů použitím vhodného matematického programu pro vykreslení grafů funkcí.)

logistický trend, logistika	$y_t = \dfrac{1}{k + a_0 a_1^{\ t}}$, nebo $\dfrac{1}{y_t} = k + a_0 a_1^{\ t}$	Křivka má tři úseky, první je charakterizován pozvolným vzestupem, druhá v okolí inflexního bodu prudkým růstem a třetí určitou vrcholovou stagnací (nasycením). Uvedený tvar je jeden z mnoha různých funkčních předpisů popisujících křivku s charakteristickým průběhem ve tvaru písmena *S*.
Gompertzova křivka	$y_t = k a_0^{\ a_1^t}$	Křivka s podobným esovitým průběhem jako logistika, ale na rozdíl od ní je asymetrická. Těžiště hodnot je až za inflexním bodem.

První tři jmenované jsou v regresní analýze běžně užívané, přičemž u exponenciály se standardně přistupuje k linearizaci logaritmováním funkčního předpisu, což získanou exponenciálu poněkud degraduje. Numerickými metodami, např. užitím *řešitele* v excelu se ale dá principu metody nejmenších čtverců vyhovět přímo, jak jsme viděli v příkladě, na který jsme se už odvolávali v 9. kapitole.

V ostatních případech už linearizace není možná. K odhadu koeficientů trendových funkcí se používá různých chytrých algoritmů, které většinou byly vymyšleny v předpočítačové éře, kdy představovaly jedinou šanci aspoň nějakého odhadu dosáhnout. Dnes se dají tyto metody využít pro určení kvalifikovaných výchozích hodnot pro nejrůznější numerické metody. (Blíže viz Seget.) (ukázka odhadu parametrů modifikované exponenciály a logistické křivky)

Analýza sezónní složky se často provádí až po očištění dat od trendové složky.
V podstatě při ní jde o určení časového úseku, po jehož uplynutí mají data zase stejnou hodnotu, příp. ovlivněnou trendovou a náhodnou složkou.
Pro studium sezónní složky se používá několika typů modelů (viz Seget). V ekonomických modelech bývá zpravidla zřejmá velikost periody (čtvtletí, měsíc), v jiných případech je nutno i tuto délku odhadovat (v hydrogeologii např. u výšky hladiny spodních vod). Používá se tu i harmonické analýzy, která modeluje průběh dat pomocí několika členů Fourierovy řady.
Parametry se určují použitím numerických metod.

Výsledků analýzy časových řad a obecně i regresní analýzy vůbec se využívá k nalezení údajů, pro které není k dispozici výsledek měření nebo pozorování. Pokud jde o chybějící údaj závislé veličiny *y* pro některou hodnotu *x* uvnitř intervalu známých hodnot *x*, jde o **interpolaci**. Ta zpravidla vede k dobrým výsledkům a nepřináší velká rizika chyb odhadované veličiny *y*.

Pokud však je nutno odhadnout výsledek *y* pro údaj *x* vně intervalu experimentálně udaných hodnot *x*, jde o **extrapolaci**. V tomto případě je nutno být opatrný, neboť matematické prostředky použité pro určení charakteru regresní závislosti nemohou zpravidla zodpovědně odhadnout budoucí nebo minulý vývoj. Uvědomte si např., že třeba rostoucí oblouk křivky třetího stupně může velmi dobře popisovat nějakou závislost, za uvažovaným intervalem hodnot *x* však může dojít k nežádoucímu propadu této kubické křivky do lokálního minima.

11. INDUKTIVNÍ STATISTIKA

Průvodce studiem

Navážeme na kapitolu 7 a ukážeme, jak pracovat se soubory, jejichž všechny prvky nejsou známy.

Předpokládané znalosti

Pojmy z předchozích kapitol, především pak ze 7. kapitoly.

Cíle

Cílem této kapitoly je vysvětlit základní pojmy statistické indukce, způsoby výběru ze základního souboru a možnosti odhadování parametrů základního souboru.

Výklad

11.1. Základní pojmy matematické statistiky a statistické indukce

Pokud jsme dosud hovořili o statistických souborech, měli jsme v souladu s definicí v 7. kapitole na mysli soubory konečného počtu prvků, u nichž jsme znali hodnotu (hodnoty) statistického znaku. Pro ně jsme pak vytvořili soustavu charakteristik, které soubor popsaly. To bylo obsahem deskriptivní statistiky.

Hlavní síla statistiky se však projeví až při práci se soubory, jejichž všechny prvky nejsou známy. Buď je jich tolik, že je prakticky nemožné (a neefektivní, finančně náročné atd.) všechny údaje o prvcích si obstarat, nebo by to třeba šlo, ale statistický soubor by tím byl zničen (např. při destrukčních zkouškách výrobků). Zavádíme tu pojem základní soubor.

Definice 11.1.1.

Základní soubor, populace (ZS)

je konečný nebo nekonečný soubor všech možných (teoreticky dosažitelných) hodnot náhodné veličiny. Hodnoty v diskrétním případě a intervaly hodnot ve spojitém případě se vyskytují ve shodě s určitým rozdělením pravděpodobnosti náhodné veličiny.

Je zřejmé, že o základním souboru v tomto smyslu nemáme úplnou informaci, ať už jde o soubory **reálné** (prvky souboru existují a teoreticky by se daly zkoumat) nebo **hypotetické** (prvky by vznikly opakováním pokusu). Ale právě o informaci o ZS stojíme, neboť jde např. o informaci o kvalitě výroby, která daným technologickým procesem vzniká apod. Tuto informaci získáváme provedením výběru ze základního souboru. Nejvhodnější by byl samozřejmě výběr, který by co nejlépe charakterizoval ZS, tj. **reprezentativní výběr**. To bychom ale museli znát vlastnosti ZS, což nebývá často. Proto vytváříme **náhodný výběr**.

11.1.1. Prostý náhodný výběr

- jedná se o pravděpodobnostní výběr, kdy každý prvek ZS (populace) má stejnou pravděpodobnost, že se do výběru dostane.

Prostý náhodný výběr lze také definovat jako výběr o rozsahu n, kdy každá množina n prvků má stejnou pravděpodobnost, že bude vybrána.

K realizaci takového výběru musíme mít k dispozici očíslovaný seznam všech prvků základního souboru - tzv. **oporu výběru**, a dále generátor náhodných čísel, pomocí něhož vybereme očíslovaný prvek z opory výběru. Předpokládejme, že ZS má N prvků a výběr bude mít n prvků. Procedura výběru sestává z následujících kroků:

1. sestavíme oporu výběru a každému prvku přiřadíme celé číslo od 1 do N
2. rozhodneme, jak velký bude rozsah výběru n
3. vygenerujeme n náhodných celých čísel mezi 1 a N
4. získáme data od prvků identifikovaných v opoře výběru těmito náhodnými čísly

Poměr mezi rozsahem výběru n a velikostí ZS (populace) N nazýváme výběrový poměr:

$$\textit{výběrový poměr} = \frac{\textit{rozsah výběru } n}{\textit{velikost populace } N}$$

Tento poměr vyjadřuje pravděpodobnost, že prvek ZS je zařazen do výběru. Výběr můžeme provádět *s vracením* nebo *bez vracení*. Vrátíme-li prvek do základního souboru, má nenulovou pravděpodobnost, že bude do výběru vybrán vícekrát. Výhodnější pro statistické odvozování různých formulí je výběr s vracením. V takovém případě je však vhodné, aby výběrový poměr byl malý (<5%).

Někdy se stává, že prostý náhodný výběr je neproveditelný nebo nákladný, hlavně v případech, kdy je ZS značně rozsáhlý. Uvádíme některé přijatelné náhradní metody výběru, jež ve výběru používají náhodný mechanismus:

- **stratifikovaný náhodný výběr** - je-li možné ZS rozdělit do dílčích oblastí, můžeme provést náhodný výběr pro každou oblast. Tyto oblasti se pak nazývají strata nebo vrstvy. Tato technika je vhodná například, když v populaci lze stratifikovat podle pohlaví, věku, ... a výzkumník chce zajistit reprezentaci každé podskupiny;

- **systematický výběr** - ze seřazeného ZS vybereme z prvních *k* prvků náhodně jeden prvek a od něho počítajíc vybereme *k*-tý, 2*k*-tý, ... prvek (viz. příklad 11.1.1.);

- **vícestupňový shlukový výběr** - často se používá pro získávání informací o veřejném mínění. Chceme například zjistit názory lidí z panelových sídlišť měst určité velikosti. Postup bude takový: 1.náhodně vybereme vzorek okresů; 2.z každého vybraného okresu se náhodně vybere určitý počet měst požadované velikosti; 3.pro tato města se náhodně vybere vzorek jejich sídlišť; 4.z vybraných sídlišť se náhodně vyberou domácnosti, ve kterých se provede dotazování. Tato vícestupňová procedura vypadá komplikovaně, ale ve skutečnosti je velmi efektivní a méně nákladná než prostý náhodný výběr domácností ze sídlišť.

Řešené úlohy

Příklad 11.1.1. Vedení vysoké školy chce provést výběr o rozsahu 50 z 1000 studentů 1.ročníku jedné z fakult, aby zjistilo spokojenost studentů s výukou matematiky.

Řešení: Může zvolit např. tuto strategii:

Jednotlivé studenty v seznamu označí čísly od 1 do 20 tak, že je v seznamu postupně očíslují touto sérií číslic jejím opakovaným použitím. Náhodně se vybere celé číslo z intervalu 1 až 20. Pak se dotáže všech studentů s tímto označením.

Jedná se tedy o systematický výběr, který je založen na pravděpodobnosti, ale prostřednictvím jiného mechanismu, než je tomu u prostého náhodného výběru.

11.2. Odhady parametrů základního souboru

Citujme nyní podrobněji ČSN 01 0250, z níž jsme již převzali předešlou definici 11.1.1.:

	Statistický soubor	**Základní soubor**	**Náhodný výběr**
Vymezení	Konečný soubor náhodné veličiny, bez vztahu k jejímu rozdělení pravděpodobnosti	Konečný nebo nekonečný soubor všech možných (teoreticky dosažitelných) hodnot náhodné veličiny. Hodnoty v diskrétním případě a intervaly hodnot ve spojitém případě se vyskytují ve shodě s určitým rozdělením pravděpodobnosti náhodné veličiny.	Konečný soubor hodnot náhodné veličiny reprezentující základní soubor. Hodnoty jsou vybrány nezávisle na sobě a hodnoty prakticky dosažitelné mají všechny stejnou možnost dostat se do výběru.
Charakterizující údaje	Ukazatelé statistického souboru charakterizují přesně a úplně vlastnosti statistického souboru. Lze je zjistit vždy ze znalosti hodnot souboru.	**Parametry** základního souboru charakterizují přesně a úplně vlastnosti základního souboru. V praxi jsou jen zřídka přesně známy, je nutno je odhadovat pomocí výběrových charakteristik.	**Charakteristiky** náhodného výběru charakterizují přibližně parametry základního souboru.
Údaje o poloze	Průměr statistického souboru (aritmetický průměr) $\overline{X} = \frac{1}{n}.\sum_{i=1}^{n} x_i$	Střední hodnota základního souboru $E(\xi) = \sum_{i=1}^{n} x_i P(x_i)$ $E(\xi) = \int_a^b x.f(x)\,dx$	Výběrový průměr $\overline{x} = \frac{1}{n}.\sum_{i=1}^{n} x_i$ Formálně platí $\overline{X} = \overline{x}$
Údaje o rozptýlení	Rozptyl statistického souboru $S^2 = \frac{1}{n}\sum_i (x_i - \overline{X})^2$	Rozptyl základního souboru $D(\xi) = \sum_{i=1}^{n} (x_i - E(\xi))^2 P(x_i)$ (diskrétní náhodná veličina), $D(\xi) = \int_a^b (x - E(\xi))^2 .f(x)\,dx$ (spojitá náhodná veličina).	Výběrový rozptyl $s^2 = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2$ Formálně platí $s^2 = \frac{n}{n-1} S^2$.

(Pozn.: Označení veličin jsme přizpůsobili označení zavedenému výše.)

V dalším textu budeme charakteristiky základního souboru (teoretické charakteristiky) značit malými písmeny, například μ, σ^2, ρ, ...

Charakteristiky empirického výběru (empirické charakteristiky), tj. charakteristiky konkrétního náhodného výběru, budeme značit malými latinskými písmeny, například m, s^2, r, ...

Výběrové charakteristiky, tj. charakteristiky obecného náhodného výběru, budeme značit velkými latinskými písmeny, například M, S^2, R, ...

Je zřejmé, že parametry základního souboru jsou konstanty, nenáhodné veličiny (které třeba ani neznáme, neboť základní soubor je možná nedostupný statistickému zpracování, popř. vůbec neexistuje), ale veličiny v posledním sloupci náhodné veličiny jsou. Mění se výběr od výběru, mění se změnou rozsahu výběru, jsou to tzv. **statistiky**. V tomto případě jsou to **bodové odhady** dvou základních parametrů základního souboru.

Definice 11.2.1.

Bodový odhad (estimátor) parametru β

je statistika B, která aproximuje parametr β s předepsanou přesností.

Oba vzorce pro bodové odhady střední hodnoty a rozptylu (viz. v tabulce výše):

$\bar{x} = \frac{1}{n}.\sum_{i=1}^{n} x_i$, $s^2 = \frac{1}{n-1}\sum_i \left(x_i - \bar{x}\right)^2$ se dají odvodit z požadavku, aby udávaly

nevychýlené odhady příslušných parametrů:

Definice 11.2.2.

Nevychýlený odhad parametru β

je taková statistika β_n, jejíž očekávaná hodnota

$E(\beta_n) = \beta$,

čili je to každá statistika, která statisticky (stochasticky) konverguje k parametru β
V opačném případě se veličina β_n nazývá odhadem vychýleným, a to vpravo nebo vlevo, podle toho, zda $E(\beta_n) - \beta > 0$, resp. $E(\beta_n) - \beta < 0$

V obou případech bodových odhadů střední hodnoty a rozptylu je také splněn požadavek **konzistentnosti (nespornosti) odhadu**:

Definice 11.2.3.

Konzistentní (nesporný) odhad parametru β

je taková statistika β_n, že pro n dosti velká je

$P(\ \beta_n - \beta\ \leq \varepsilon) > 1 - \eta$,

kde $\varepsilon > 0$, $\eta > 0$ jsou jakákoliv (libovolně malá) předem zvolená čísla.

K získávání bodových odhadů se používají dvě metody:

a) **metoda momentů**

je založena na porovnání momentů základního souboru a výběru. Počet prorvnávaných momentů je dán počtem parametrů rozdělení. Závisí-li rozdělení na S – parametrech, řešíme soustavu S rovnic o S neznámých:

$$\begin{aligned} \mu_1 &= m_1 \\ \mu_2 &= m_2 \\ &\vdots \\ \mu_S &= m_S \end{aligned}$$

μ_i … teoretické momenty, m_i … empirické momenty; $i = 1,2,\dots,S$

Řešené úlohy

Příklad 11.2.1. Metodou momentů určete neznámý parametr Poissonova rozdělení.

Řešení: Poissonovo rozdělení má pravděpodobnostní funkci:

$$p\ \ x,\lambda\ \ = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

Vybereme n prvků $x_1, \dots, x_n$

$$\mu_1 = \lambda$$

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\mu_1 = m_1$$

Tedy:

$$\lambda = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

Řešené úlohy

Příklad 11.2.2. Metodou momentů určete neznámý parametr exponenciálního rozdělení.

Řešení: Exponenciální rozdělení má hustotu pravděpodobnosti:

$$f(x) = \begin{cases} 0 & x < 0 \\ \lambda \cdot e^{-\lambda x} & x \geq 0 \end{cases}$$

Vybereme n prvků $x_1, \ldots, x_n$

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\mu_1 = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda x} dx = \lambda \cdot \int_0^{\infty} x \cdot e^{-\lambda x} dx = \begin{vmatrix} u = x & v' = e^{-\lambda x} \\ u' = 1 & v = -\frac{1}{\lambda} \cdot e^{-\lambda x} \end{vmatrix} =$$

$$= \left[-x \cdot e^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \lim_{x \to \infty} \frac{-x}{e^{\lambda x}} + 0 - \left[\frac{1}{\lambda} \cdot e^{-\lambda x}\right]_0^{\infty} = 0 + \frac{1}{\lambda} = \frac{1}{\lambda}$$

Porovnáme-li tedy opět první počáteční momenty:

$$\mu_1 = m_1$$

$$\frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

b) **metoda maximální věrohodnosti**

Má-li základní soubor frekvenční funkci $p(x,\theta)$, kde $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ jsou parametry rozdělení základního souboru, pak pravděpodobnost, že výběr $\xi_1, \xi_2, \ldots, \xi_n$ bude mít realizaci $x_1, x_2, \ldots, x_n$ je vyjádřena vztahem:

$$P(\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_n = x_n) = p(x_1,\theta).p(x_2,\theta)\cdot \ldots \cdot p(x_n,\theta) = \prod_{i=1}^{n} p(x_i,\theta) =$$

$$= L(x_1, x_2, \ldots, x_n, \theta)$$

Funkci L nazýváme funkcí maximální věrohodnosti.

Za nejpravděpodobnější považujeme takovou hodnotu θ, při níž má funkce L maximální hodnotu.

Řešené úlohy

Příklad 11.2.3. Metodou maximální věrohodnosti odhadněte neznámý parametr Poissonova rozdělení.

Řešení: Poissonovo rozdělení má pravděpodobnostní funkci:

$$p(x,\lambda) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

$$L(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda} \qquad |\ln$$

$$\ln L = \sum_{i=1}^{n} \left(\ln \lambda^{x_i} - \ln(x_i!) - \lambda \right)$$

$$\ln L = \sum_{i=1}^{n} \left(x_i \cdot \ln \lambda - \ln(x_i!) - \lambda \right)$$

$$\frac{d \ln L}{d\lambda} = \sum_{i=1}^{n} \left(x_i \cdot \frac{1}{\lambda} - 1 \right)$$

Položíme-li derivaci rovnu 0:

$$\frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0$$

$$\frac{1}{\lambda} \sum_{i=1}^{n} x_i = n$$

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Kritické hodnoty rozdělení

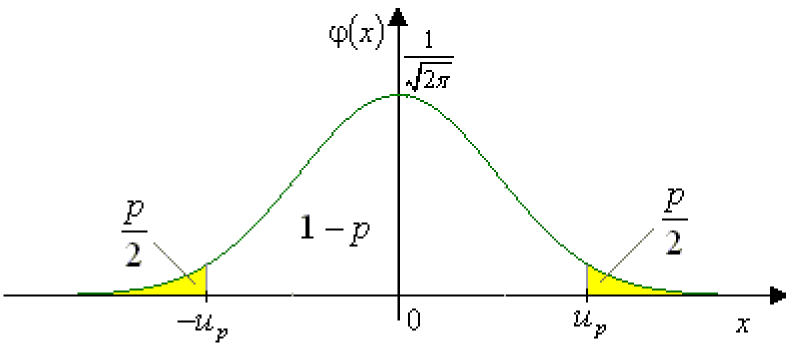
Definice 11.2.4.

Kritické hodnoty rozdělení na hladině významnosti p jsou kvantily, kde index p vyjadřuje pravděpodobnost, že náhodná veličina (u symetrických rozdělení její absolutní hodnota), překročí tuto hodnotu.

Užívaná označení:

u_p – kritická hodnota normálního rozdělení na hladině významnosti p.

$P(|X| > u_p) = p$, X ...má normované normální rozdělení $N(0,1)$



$$\Phi(u_p) - \Phi(-u_p) = 1 - p$$

$$\Phi(u_p) - \left[1 - \Phi(u_p)\right] = 1 - p$$

$$2\Phi(u_p) = 2 - p$$

$$\Phi(u_p) = 1 - \frac{p}{2}$$

, kde $u_p \ldots \left(1 - \frac{p}{2}\right)$-kvantil normálního rozdělení $N(0,1)$

Odsud se určí např. $u_{0,05} = 1,96$.

$\chi^2_{p(n)}$ – kritická hodnota rozdělení χ^2 s n-stupni volnosti na hladině významnosti p.

$P(X > \chi^2_{p(n)}) = p$, X ...má rozdělení χ^2 s n-stupni volnosti

$t_{p(n)}$– kritická hodnota Studentova rozdělení s n-stupni volnosti na hladině významnosti p.

$P(|X| > t_{p(n)}) = p$, X ...má Studentovo rozdělení s n-stupni volnosti

$F_{p(m,n)}$– kritická hodnota Fischerova rozdělení s m,n-stupni volnosti na hladině významnosti p.

$P(X > F_{p(m,n)}) = p$, X ...má Fischerovo rozdělení s m,n-stupni volnosti

Intervalové odhady parametrů:

Definice 11.2.4.

Intervalový odhad parametru β základního souboru

je interval $< B_1 ; B_2 >$, v němž leží skutečná hodnota parametru s pravděpodobností 1 - p, tzn.

$P(B_1 \leq \beta \leq B_2) = 1 - p$.

Interval $< B_1 ; B_2 >$ se nazývá interval spolehlivosti (konfidenční interval) pro parametr β na hladině významnosti $\boldsymbol{p}$ (nebo se stupněm spolehlivosti 1 - $\boldsymbol{p}$).

Hodnoty B_1, B_2 jsou kritické hodnoty pro parametr β.

Intervaly ($-\infty$; B_1) a (B_2 ; $+\infty$) se nazývají kritické intervaly.

Hladina významnosti p je pravděpodobnost toho, že skutečná hodnota odhadovaného parametru **neleží** uvnitř intervalu spolehlivosti. Bývá zvykem volit hodnotu $p = 0{,}1$ nebo $p = 0{,}05$ nebo $p = 0{,}01$.

Stupeň spolehlivosti vyjadřuje pravděpodobnost toho, že skutečná hodnota parametru **leží** v intervalu spolehlivosti.

Interval spolehlivosti lze určit nekonečně mnoha způsoby. Nejčastěji se používá **symetrický oboustranný interval spolehlivosti**, tzn. že parametr β se vyskytuje v jednom z kritických intervalů s pravděpodobností $\frac{p}{2}$.

$$P(\beta < B_1) = P(\beta > B_2) = \tfrac{p}{2}.$$

Věnujme se nyní intervalovému odhadu nejdůležitějších statistických veličin, střední hodnoty a rozptylu. Ukazuje se, že ten se dá odvodit jako důsledek tzv. **centrální limitní věty**. Uveďme ji v jednom z několika užívaných tvarů bez důkazu:

Věta 11.2.1.

Nechť $X = X_1 + X_2 + \ldots + X_n$ je náhodná veličina, která vznikla součtem nezávislých náhodných veličin s konečnou střední hodnotou μ a konečným rozptylem σ^2.

Pak náhodná proměnná $Y_n = \dfrac{\frac{X_1 + X_2 + \cdots X_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}$ má pro $n \to \infty$ normální rozložení $N(0,1)$.

Všimněme si hlavně toho, že o výchozím (základním) souboru není předpokládáno s výjimkou konečnosti jeho základních charakteristik vůbec nic. Hlavně se nic nepředpokládá o jeho rozložení. Přesto je tedy dokazatelné, že výběrové průměry normální rozložení mají. A jejich střední hodnota je rovna střední hodnotě základního souboru (vzpomeňme na bodový odhad střední hodnoty) a rozptyl těchto průměrů je *n-tinou* rozptylu základního souboru.

Zde si můžete otevřít **ilustrační úlohu** vyřešenou v Excelu (pouze na webu).

11.2.1. Intervalový odhad střední hodnoty

Jsou-li tedy u_1, u_2 dva libovolné kvantily normovaného normálního rozložení, platí

$$P\left(u_1 \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq u_2\right) = \int_{u_1}^{u_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

Protože však nejčastěji volíme konfidenční interval, do něhož má s předem danou pravděpodobností padnout střední hodnota základního souboru, souměrný kolem bodového odhadu, upravujeme vzorec pro **intervalový odhad střední hodnoty** do tvaru:

$$\left|\overline{x}-\mu\right| < u_p.\frac{\sigma}{\sqrt{n}}, \text{ čili } P\left(\mu \in \left(\overline{x}-u_p.\frac{\sigma}{\sqrt{n}}; \overline{x}+u_p.\frac{\sigma}{\sqrt{n}}\right)\right) = 1-p.$$

Přitom jsme písmenem p označili **hladinu významnosti**, u_p **příslušný kvantil normovaného normálního rozložení** . Hodnota **1 -** p je pak **hladina spolehlivosti** (např. pro p = 0,05 je $u_{0,05}$ = 1,96). Výrazem $\overline{x}$ jsme označili bodový odhad střední hodnoty, jak je běžně zvykem.

Pokud není známa hodnota rozptylu základního souboru σ (tak je tomu většinou), nahradíme ji bodovým odhadem. Podmínce asymptotičnosti ovšem nutno vyhovět a užívat vzorec pouze pro $n > 30$.

Pro **menší vzorky** platí analogický vztah, ale normální normované rozložení je nahrazeno **rozložením Studentovým s n - 1 stupni volnosti**. Kvantil u_p pak nahrazujeme kvantilem t_p (n-1) Studentova t-rozložení.

(**Počet stupňů volnosti**, který teď bude u některých speciálních rozložení pravděpodobnosti vystupovat, bude vždy označovat **počet nezávislých pozorování**, která jsou v dané situaci volitelná. Tak v případě odhadu střední hodnoty je možno u vzorku o rozsahu *n* zvolit *n* - 1 hodnot libovolně, n-tý prvek je z dané střední hodnoty dopočitatelný. Odečetl se tedy jeden stupeň volnosti, neboť existovala jedna vazba mezi uvažovanými veličinami. Analogický postup se pro výpočet stupňů volnosti užívá obecně.)

Výraz $\Delta = u_p.\frac{\sigma}{\sqrt{n}} = u_p.\frac{s}{\sqrt{n-1}}$, resp. $\Delta = t_p.\frac{\sigma}{\sqrt{n}} = t_p.\frac{s}{\sqrt{n-1}}$ je vlastně **požadovaná přesnost** pro hledaný parametr (běžný je zápis $\mu = \bar{x} \pm \Delta$), která platí pro zvolenou hladinu významnosti p. Ze vztahu pro výpočet Δ však můžeme naopak určit n, které určí potřebný rozsah výběru, jehož charakteristika má požadovanou spolehlivost:

$$n = \left(\frac{u_p.\sigma}{\Delta}\right)^2, \text{ resp. } n = 1 + \left(\frac{s.t_p}{\Delta}\right)^2$$

Bez problémů je tato inverzní úloha pro případ, že používáme předpoklad o normalitě. Při aplikaci Studentova *t*-rozložení se vyskytuje hledané n na obou stranách rovnice v implicitní podobě.

Řešené úlohy

Příklad 11.2.1. Měřili jsme průměr vačkového hřídele na 250 součástkách. Předpokládáme normální rozdělení souboru. Z výsledků měření jsme určili výběrový průměr a výběrovou disperzi x_p = 995,6, s^2 = 134,7. Určete interval spolehlivosti pro střední hodnotu základného souboru při hladině významnosti 5 %.

Řešení: Úlohu vyřešíme v Excelu - z důvodu jednoduchého výpočtu kritické hodnoty normálního rozdělení pomocí předdefinované funkce NORMSINV - v souladu s předchozí teorií:

$$\Delta = \frac{s}{\sqrt{n-1}}.u_p = \frac{\sqrt{134{,}7}}{\sqrt{249}}.\text{NORMSINV}\ \ 0{,}975\ \ = 1{,}441558$$

Intervalový odhad střední hodnoty je tedy:

$$\langle x_p - \Delta; x_p + \Delta \rangle = \langle 994{,}1584; 997{,}0416 \rangle$$

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Příklad 11.2.2. Při měření kapacity sady kondenzátorů bylo provedeno 10 měření s výsledky v tabulce. Odhadněte interval spolehlivosti pro kapacitu těchto kondenzátorů se spolehlivostí 90 %, resp. 95 %.

152	156	148	153	150	156	140	155	145	148

Řešení: Úlohu vyřešíme obdobně jako předchozí příklad 11.2.1.:

Výběrový průměr x_p a výběrovou směrodatnou odchylku s vypočteme v Excelu pomocí předdefinovaných funkcí PRŮMĚR a SMODCH. Výsledky:

$x_p = 150{,}3$; $s = 4{,}92$

$$\Delta_{0,90} = \frac{s}{\sqrt{n-1}} . t_p\left(n-1\right) = \frac{4{,}92}{\sqrt{9}} .\mathrm{TINV}\left(0{,}1;9\right) \doteq 3{,}0065$$

$$\Delta_{0,95} = \frac{s}{\sqrt{n-1}} . t_p\left(n-1\right) = \frac{4{,}92}{\sqrt{9}} .\mathrm{TINV}\left(0{,}05;9\right) \doteq 3{,}7102$$

Interval spolehlivosti na hladině významnosti 90%:

$$\left\langle x_p - \Delta; x_p + \Delta \right\rangle = \left\langle 147{,}29; 153{,}31 \right\rangle$$

Interval spolehlivosti na hladině významnosti 95%:

$$\left\langle x_p - \Delta; x_p + \Delta \right\rangle = \left\langle 146{,}59; 154{,}01 \right\rangle$$

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

11.2.2. Intervalový odhad rozptylu

Přistupme nyní k odvození intervalového odhadu disperze. V 5. kapitole o rozloženích pravděpodobnosti spojité náhodné veličiny bylo konstatováno, že náhodná veličina, která vznikne součtem normovaných veličin s normálním rozložením, má Pearsonovo rozložení χ^2. Stejně tak často tuto součtovou veličinu i označujeme, tedy

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{\sigma^2}$$ má rozložení χ^2 s n stupni volnosti.

Neznáme-li střední hodnotu (a to zpravidla platí), pak náhodná veličina $\chi^2 = \sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{\sigma^2}$ má Pearsonovo rozložení pro (n - 1) stupňů volnosti.

Dvoustranný intervalový odhad náhodné veličiny χ^2 můžeme zapsat pravděpodobnostní rovnicí:

$$P\left(\chi^2_{1-\frac{p}{2}}(n-1) \le \chi^2 \le \chi^2_{\frac{p}{2}}(n-1)\right) = 1-p \text{ čili } P\left(\chi^2_{1-\frac{p}{2}}(n-1) \le \frac{n.S^2}{\sigma^2} \le \chi^2_{\frac{p}{2}}(n-1)\right) = 1-p\,.$$

Kritické hodnoty jsou tabelovány.

Po úpravě získáme pravděpodobnostní rovnici pro **intervalový odhad rozptylu** základního souboru v praktičtějším tvaru:

$$P\left(\frac{n.S^2}{\chi^2_{\frac{p}{2}}(n-1)} \le \sigma^2 \le \frac{n.S^2}{\chi^2_{1-\frac{p}{2}}(n-1)}\right) = 1-p$$

Řešené úlohy

Příklad 11.2.3. Určete oboustranný konfidenční interval rozptylu normálně rozloženého základního souboru pro hladiny spolehlivosti 0,90, 0,95 a 0,99, když u výběru s rozsahem n = 12 byl zjištěn rozptyl 0,64. Posuďte získané výsledky.

Řešení: Kritické hodnoty Pearsonova rozdělení v excelu vypočteme pomocí předdefinované funkce CHIINV.

Řešení pro spolehlivost 0,90:

$$\frac{n.s^2}{\chi^2_{\frac{p}{2}}(n-1)} \le \sigma^2 \le \frac{n.s^2}{\chi^2_{1-\frac{p}{2}}(n-1)}$$

$$\frac{12.0,64}{\text{CHIINV}(0,05;11)} \le \sigma^2 \le \frac{12.0,64}{\text{CHIINV}(0,95;11)}$$

$$0,358 \le \sigma^2 \le 1,539$$

Zbývající dva případy vyřešíme zcela analogicky.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Úlohy k samostatnému řešení

11.1. Měřil se průměr hřídele na 250 součástkách. Předpokládáme normální rozdělení souboru. Z výsledků se určil výběrový průměr a výběrová disperze: $\bar{x}$ = 995,6; s^2 = 134,7. Určete interval spolehlivosti pro střední hodnotu na hladině významnosti 5%.

11.2. Byla měřena délka trvání určitého procesu. Z 12 měření byla zjištěna střední doba trvání procesu 44 s a směrodatná odchylka 4 s. Sestrojte 90 % a 95 % interval spolehlivosti pro očekávanou délku procesu za předpokladu normálního rozdělení.

11.3. Při měření kapacity sady kondenzátorů bylo provedeno 10 měření s výsledky:
152, 156, 148, 153, 150, 156, 140, 155, 145, 148.
Odhadněte interval spolehlivosti pro kapacitu těchto kondenzátorů se spolehlivostí *a*) 90%, *b*) 95%.

11.4. Bylo zkoušeno 30 náhodně vybraných ocelových tyčí k určení meze kluzu určitého druhu oceli. Po zpracování výsledků byla určena její empirická střední hodnota 286,4 Mpa a rozptyl 121 [Mpa^2]. Určete intervalový odhad parametrů základního souboru s 95% spolehlivostí. Kolik vzorků by bylo třeba volit, aby chyba určené střední hodnoty nepřesáhla 2 Mpa?

11.5. Určete intervalový odhad s 90% spolehlivostí střední hodnoty a směrodatné odchylky pro následující hodnoty:
606, 1249, 267, 44, 510, 340, 109, 1957, 463, 801, 1086, 169, 233, 1734, 1458, 80, 1023, 2736, 917, 459.

Výsledky úloh k samostatnému řešení

11.1. <994,16;997,04>

11.2. p = 0,1: <41,83;46,17>

p = 0,05: <41,35;46,65>

11.3. *a*) <147,29;153,31>

b) <146,59;154,01>

11.4. <282,22;290,58>

<79,39;226,21>

n = 120

11.5. <544,24;1101,55>

<572,22;987,73>

12. TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

Průvodce studiem

Navážeme na předchozí kapitolu 11 a vysvětlíme některé statistické testy.

Předpokládané znalosti

Pojmy z předchozích kapitol.

Cíle

Cílem této kapitoly je vysvětlit postup při testování statistických hypotéz a seznámit s některými konkrétními statistickými testy.

Výklad

12.1. Statistické hypotézy - úvod

Od statistického šetření neočekáváme pouze elementární informaci o velikosti některých statistických ukazatelů. Používáme je i k ověřování našich očekávání o výsledcích nějakého procesu, k posuzování významnosti změn, které byly způsobeny změnou technologie, apod. Ukážeme, že ač formulace úloh toho typu se liší od formulace úlohy o odhadech parametrů, jde zpravidla vždy o řešení inverzní úlohy o intervalovém odhadu. Zaveďme si však napřed příslušnou terminologii.

Definice 12.1.1.

Statistická hypotéza

je tvrzení, které se týká neznámé vlastnosti rozdělení pravděpodobnosti náhodné proměnné (i vícerozměrné) nebo jejích parametrů.

Hypotéza, jejíž platnost ověřujeme, se nazývá nulová hypotéza $\boldsymbol{H}_0$.

Proti nulové hypotéze stavíme alternativní hypotézu $\boldsymbol{H}_1$. Ta může být buď oboustranná nebo jednostranná. Pak i testy jsou buď oboustranné nebo jednostranné.

Hypotézy se mohou týkat pouze neznámých číselných parametrů rozložení náhodné veličiny, pak jde o testy parametrické.

Ostatní typy jsou testy neparametrické.

Statistické testy

jsou postupy, jimiž prověřujeme platnost nulové hypotézy. Na základě nich pak hypotézu buď přijmeme nebo odmítneme.

Testovací kritérium

je náhodná veličina závislá na náhodném výběru (též nazývaná statistika) mající vztah k nulové hypotéze.

Jednostranné a oboustranné testy se od sebe rozlišují z hlediska alternativní hypotézy, kterou stavíme proti prověřované nulové hypotéze a která může být dvojího druhu, jak plyne z tohoto příkladu:

Nechť nulová hypotéza předpokládá, že $\boldsymbol{A} = \boldsymbol{B}$. V případě, že tuto hypotézu zamítneme, je buď $\boldsymbol{A} \neq \boldsymbol{B}$, nebo $\boldsymbol{A} > \boldsymbol{B}$ (resp. $\boldsymbol{A} < \boldsymbol{B}$).

a) V prvém případě ($\boldsymbol{A} \neq \boldsymbol{B}$) nebereme zřetel na znaménko rozdílu $\boldsymbol{A} - \boldsymbol{B}$, takže může být buď $\boldsymbol{A} - \boldsymbol{B} < \mathbf{0}$ nebo $\boldsymbol{A} - \boldsymbol{B} > \mathbf{0}$. V těchto případech používáme **oboustranný test**.

b) V druhém případě, kdy proti hypotéze $\boldsymbol{A} = \boldsymbol{B}$ klademe možnost $\boldsymbol{A} > \boldsymbol{B}$ (resp. $\boldsymbol{A} < \boldsymbol{B}$), používáme **jednostranných testů**.

Pro **kritické hodnoty** testovacího kritéria a_p, b_p platí:

$$P\left(a_p \leq X \leq b_p\right) = 1 - p.$$

Tyto hodnoty oddělují **interval prakticky možných hodnot** (interval spolehlivosti, konfidenční interval) $\langle a_p, b_p \rangle$ od **kritických intervalů**, v nichž se hodnoty veličiny X vyskytují s pravděpodobností p, které říkáme hladina významnosti. Nejčastěji volíme $p = 0{,}01$ nebo $p = 0{,}05$.

Pro oboustranné odhady volíme:

$$P(X < a_p) = P(X > b_p) = \frac{p}{2},$$

pro jednostranné buď

$$P(X < a_p) = 0, \quad P(X > b_p) = p \text{ nebo}$$

$$P(X < a_p) = p, \quad P(X > b_p) = 0.$$

Porovnání hodnoty testovacího kritéria s jeho kritickými hodnotami slouží k rozhodnutí o výsledku testu. Musíme si uvědomit, že nemůžeme mluvit o **dokazování** správnosti či nesprávnosti zvolené hypotézy - to není v možnostech statistické indukce. Závěr testu pouze rozhodne mezi dvěmi možnostmi:

- **hypotézu přijímáme** (zamítáme alternativní hypotézu), leží-li pozorovaná hodnota testovacího kritéria v intervalu prakticky možných hodnot. Znamená to, že rozdíl mezi pozorovanou a teoretickou hodnotou testovacího kritéria je vysvětlitelný na dané hladině významnosti p náhodností výběru.
- **hypotézu zamítáme** (přijímáme alternativní hypotézu), leží-li pozorovaná hodnota testovacího kritéria v kritickém oboru. Rozdíly považujeme za statisticky významné na zvolené hladině významnosti p, tzn., že se nedají vysvětlit pouze náhodností výběru.

Příklady otázek, na které se dá odpovídat pomocí výsledků příslušných statistických testů:

- Má základní soubor (ZS) předpokládanou střední hodnotu?
- Mají dva soubory stejnou disperzi?
- Můžeme předpokládat, že dva výběry pocházejí z téhož ZS?
- Má ZS předpokládané rozdělení?
  atd.

Těmito slovy jistě nebudou technici formulovat své otázky v konkrétním průmyslovém podniku. Bude je ale např. zajímat, zda

- bylo dodáno uhlí deklarované kvality
- dva měřící přístroje pracují stejně přesně
- se nezměnily provozní podmínky ovlivňující výrobu (např. seřízení obráběcích strojů)
- produkce zmetků v jednotlivých hodinách je rovnoměrná

(Pokuste se popsat konkrétní provozní realizace výše uvedených situací.)

Ve shodě s běžnými zvyklostmi definujme:

Definice 12.1.2.

Nechť ***b*** je pozorovaná, kdežto β teoretická hodnota statistiky ***B*** a nechť $\langle \boldsymbol{a_p}, \boldsymbol{b_p} \rangle$ je interval prakticky možných hodnot veličiny ***B*** na 100*p*% hladině významnosti.

Pak říkáme, že rozdíl ***b*** - β je

1. náhodně vysvětlitelný, když $b \in \langle a_{0,05}; b_{0,05} \rangle = J_{0,05}$;
2. statisticky významný, když $b \in \langle a_{0,01}; b_{0,01} \rangle = J_{0,01}$;
3. slabě statisticky významný, když $b \notin J_{0,05}$, ale $b \in J_{0,01}$.

12.1.1. Kroky při testování hypotézy

- Formulace výzkumné otázky ve formě nulové a alternativní statistické hypotézy
- Zvolení přijatelné úrovně chyby rozhodování (volba hladiny významnosti *p*)
- Volba testovacího kritéria
- Výpočet hodnoty testovacího kritéria
- Určení kritických hodnot testovacího kritéria
- Doporučení (přijmutí nebo zamítnutí nulové hypotézy H_0)

Poznámky

Hladina významnosti je pravděpodobnost, že se zamítne nulová hypotéza, ačkoliv ona platí. Pochopitelně se tato hodnota volí velmi malá, jak již bylo řečeno, nejčastěji 0,05 nebo 0,01.

Jestliže test neindikuje zamítnutí nulové hypotézy H_0*, je nesprávné přijmout nulovou hypotézu jako definitivně pravdivou. Správně můžeme pouze prohlásit, že není dostatek dokladů pro zamítnutí nulové hypotézy.*

Netvrďme, že data ukazují, že teorie platí/neplatí. Správnější je říct, že data podporují nebo nepodporují rozhodnutí o zamítnutí platnosti nulové hypotézy.

12.1.2. Test jako rozhodování

Při testování hypotéz mohou nastat čtyři možnosti, které popisuje následující tabulka:

		Závěr testu	
		H_0 platí	H_0 neplatí
Skutečnost	H_0 platí	správný	chyba I.druhu
	H_0 neplatí	chyba II.druhu	správný

Existují tedy dvě možnosti chyby:

- **chyba I. druhu** - nulová hypotéza platí, ale zamítne se;
- **chyba II. druhu** - nulová hypotéza neplatí, ale přijme se.

Přirovnáme-li tuto situaci k medicínskému testování, pak chyba I. druhu znamená falešně pozitivní výsledek (pacient je zdráv, ale testování ukazuje na nemoc), chyba II. druhu odpovídá falešně negativnímu výsledku (pacient je nemocný, ale test to neodhalí).

Pravděpodobnost chyby I. druhu je podmíněná pravděpodobnost, že zamítneme nulovou hypotézu za předpokladu, že platí - označujeme p - viz. výše. Pravděpodobnost chyby II. druhu je podmíněná pravděpodobnost, že nezamítneme nulovou hypotézu za předpokladu, že neplatí, označujeme p_0:

$$P(\text{chyba I. druhu} \mid H_0 \text{ platí}) = p$$
$$P(\text{chyba II. druhu} \mid H_1 \text{ neplatí}) = p_0$$

Konvenční hodnoty pro p_0 jsou 0,2 nebo 0,1.

Někdy můžeme také mluvit o opačných jevech k chybě I. a II. druhu, tzn. o podmíněné pravděpodobnosti, že neuděláme chybu I.druhu (spolehlivost testu) nebo že neuděláme chybu II. druhu. **Síla testu** odpovídá hodnotě (1 - p_0). Jedná se tedy o podmíněnou pravděpodobnost, že správně odhalíme testem neplatnost nulové hypotézy:

$$P(\text{neuděláme chybu I. druhu} \mid H_0 \text{ platí}) = 1 - p = \text{„spolehlivost“}$$
$$P(\text{neuděláme chybu II. druhu} \mid H_1 \text{ neplatí}) = 1 - p_0 = \text{„síla testu“}$$

Cílem při testování nulové hypotézy je omezit úrovně pravděpodobnosti chyb I. a II. druhu. Jinými slovy - usilujeme o maximalizaci spolehlivosti a síly testu.

Řešené úlohy

Příklad 12.1.1. Testování přiblížíme pomocí analogie se soudním procesem. Má padnout rozhodnutí, zda obžalovaný spáchal či nespáchal zločin.

Řešení: Soudní systém se řídí zásadou, že obžalovaný je nevinen, dokud se nepodaří prokázat opak. Formulace hypotéz má tedy tuto podobu:

H_0: Obžalovaný je nevinen.

H_1: Obžalovaný je vinen.

Různé možnosti vztahu mezi pravdou a rozhodnutím soudu vidíme v tabulce:

		Závěr soudu	
		Obžalovaný je nevinen	**Obžalovaný je vinen**
Skutečnost	Obžalovaný je nevinen	správný	chyba I. druhu
	Obžalovaný je vinen	chyba II. druhu	správný

Uvědomme si, že chyba I. druhu má pro jedince fatální následky. Proto její možnost eliminujeme na nejmenší možnou míru. Soud musí jasně prokázat vinu obžalovaného. Jeho rozhodnutí také podléhají přezkoumání vyšších instancí. Odpovídá to volbě velmi malé hladiny významnosti. V mnoha jiných případech však nevíme zcela přesně, která chyba je pro nás důležitější.

V další části uvedeme některé důležité statistické testy:

12.2. Hypotézy o rozptylu

12.2.1. Test významnosti rozdílu dvou rozptylů (F-test)

Předpoklady:

Jsou dány dva výběry o rozsazích n_1, n_2 s rozptyly S_1^2, S_2^2, vybrané ze dvou základních souborů s rozděleními $N(\mu_1; \sigma_1^2)$ a $N(\mu_2; \sigma_2^2)$.

Nulová hypotéza:

H_0: $\sigma_1^2 = \sigma_2^2$

Alternativní hypotéza:

H_1: $\sigma_1^2 \neq \sigma_2^2$

Testovací kritérium:

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{n_1\left(n_2-1\right).S_1^2}{n_2\left(n_1-1\right).S_2^2}$$

má Fisherovo-Snedecorovo rozdělení $F(n_1 - 1, n_2 - 1)$.

Závěr:

Jestliže $F > F_{\frac{p}{2}}\left(n_1-1, n_2-1\right)$, zamítáme hypotézu H_0 (přijímáme H_1).

Indexy 1, 2 volíme tak, aby testovací kritérium $F > 1$.

Poznámka

V případě, že bychom chtěli prokázat hypotézu H_0 *proti hypotéze* H_1: $\sigma_1^2 > \sigma_2^2$, *použili bychom kritickou hodnotu* $F_{\mathbf{p}}(n_1 - 1, n_2 - 1)$

Řešené úlohy

Příklad 12.2.1. Byly sledovány výsledky běhu na 50 m (v sekundách) u skupiny desetiletých chlapců a dívek. Posuďte získané výsledky z hlediska vyrovnanosti výkonů v jednotlivých skupinách.

Chlapci:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
10,80	9,30	9,40	9,90	10,20	9,30	9,40	8,90	8,90	9,60	9,70	10,60	9,40	9,50	9,60	10,00	9,30

18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
9,40	8,40	9,80	8,80	9,20	9,50	9,80	9,00	10,50	9,40	9,30	9,90	9,10	9,60	8,70	8,10

Dívky:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
10,70	10,80	10,00	10,60	9,20	10,20	9,90	10,00	9,30	10,20	9,80	10,00	10,00	11,00

15	16	17	18	19	20	21	22	23	24	25	26	27	28
12,00	10,00	10,00	11,20	9,40	10,70	9,30	10,10	9,10	10,20	9,30	10,00	9,40	10,90

Řešení: Hladinu významnosti zvolíme $p = 0{,}05$.

Určíme potřebné charakteristiky u obou skupin (prohodili jsme pořadí tak, aby vyšlo $F > 1$):

Dívky:	Chlapci:
$n_1 = 28$	$n_2 = 33$
$s_1^2 = 0{,}4521$	$s_2^2 = 0{,}3302$

Určíme hodnotu testovacího kritéria:

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{n_1\,(n_2 - 1)\,.s_1^2}{n_2\,(n_1 - 1)\,.s_2^2} = \frac{28.32.0{,}4521}{33.27.0{,}3302} \doteq 1{,}377$$

Kritická hodnota (vypočtená např. v Excelu pomocí předdefinované funkce FINV):

$F_{0{,}025}(27{,}32) = \mathrm{FINV}(0{,}025;27;32) = 2{,}0689$

Testovací kritérium nepřekročilo kritickou hodnotu, tudíž přijmeme H_0. Mezi rozptyly není statisticky významný rozdíl.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

12.3. Hypotézy o střední hodnotě

12.3.1. Test významnosti rozdílu $|M - \mu_0|$

Předpoklady:

Je dán výběr ze základního souboru s rozdělením $N(\mu; \sigma^2)$ o rozsahu n se střední hodnotou M a disperzí S^2.

Nulová hypotéza:

H_0: $\mu = \mu_0$

Alternativní hypotéza:

H_1: $\mu \neq \mu_0$

Testovací kritérium:

$$T = \frac{M - \mu_0}{S}.\sqrt{n-1}$$

má Studentovo rozdělení $t(n - 1)$.

Závěr:

Jestliže $|T| > t_p(n - 1)$, zamítáme hypotézu H_0 (přijímáme H_1).

Poznámka

Volíme-li alternativní hypotézu H_1: $\mu > \mu_0$, pak hodnotu testovacího kritéria srovnáváme s kritickou hodnotou $t_{2p}(n - 1)$.

Řešené úlohy

Příklad 12.3.1. V pivovaru došlo k opravě plnící linky. Na hladině významnosti $p = 0{,}05$ ověřte, zda se oprava zdařila, tj., zda linka plní do láhví pivo o objemu 500ml. Výsledky u vybraných vzorků (v mililitrech):

495,2	496,8	502,1	498,5	501	503	500,7
501,5	501,8	499,1	500,9	502,2	501,7	500,4
500,2	501,1	499,9	500,2	501,1	500,8	499,3

Řešení: $\mu_0 = 500$, tudíž:

H_0: $\mu = 500$

H_1: $\mu \neq 500$

Výpočet základních charakteristik:

$n = 21$ $M = 500{,}3571$ $S = 1{,}77806$

Testovací kritérium:

$$T = \frac{M - \mu_0}{S}.\sqrt{n-1} = \frac{500{,}3571 - 500}{1{,}77806}.\sqrt{20} \approx 0{,}898$$

Kritická hodnota (vypočteme např. v Excelu pomocí předdefinované funkce TINV):

$t_{0,05}(20)$ = TINV(0,05;20) = 2,086

Závěr:

Testovací kritérium nepřekročilo kritickou hodnotu, tudíž přijmeme H_0. Oprava se zdařila, linka plní lahve správně.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

12.3.2. Test významnosti rozdílu dvou výběrových průměrů (*t*-test)

Předpoklady:

Jsou dány dva výběry o rozsazích n_1, n_2 se středními hodnotami M_1, M_2 a disperzemi S_1^2, S_2^2, které pocházejí ze dvou základních souborů s rozděleními $N(\mu_1;\sigma_1^2)$ a $N(\mu_2;\sigma_2^2)$.

Nulová hypotéza:

H_0: $\mu_1 = \mu_2$

Alternativní hypotéza:

H_1: $\mu_1 \neq \mu_2$

a) jestliže můžeme předpokládat $\sigma_1^2 = \sigma_2^2$ (prověříme *F*-testem), volíme **testovací kritérium:**

$$T = \frac{M_1 - M_2}{\sqrt{n_1.S_1^2 + n_2.S_2^2}}.\sqrt{\frac{n_1.n_2.\ n_1 + n_2 - 2}{n_1 + n_2}} \ ,$$

které má Studentovo rozdělení $t(n_1 + n_2 - 2)$.

Závěr:

Jestliže $|\,T\,| > t_p$, zamítneme H_0.

b) jestliže předpokládáme $\sigma_1^2 \neq \sigma_2^2$ (prověříme *F*-testem), volíme **testovací kritérium:**

$$T = \frac{M_1 - M_2}{\sqrt{\ n_2 - 1\ .S_1^2 +\ n_1 - 1\ .S_2^2}}.\sqrt{\ n_1 - 1\ .\ n_2 - 1\ } \ ,$$

které má rozdělení, složené ze dvou Studentových rozdělení.

Kritické hodnoty určíme podle vzorce:

$$t_p = \frac{\ n_2 - 1\ .S_1^2.t_p\ \ n_1 - 1\ +\ n_1 - 1\ .S_2^2.t_p\ \ n_2 - 1\ }{\ n_2 - 1\ .S_1^2 +\ n_1 - 1\ .S_2^2}$$

Závěr:

Jestliže $|\,T\,| > t_p(n_1 + n_2 - 2)$, zamítneme H_0.

Poznámka

t-test používáme např. k ověřování následujících hypotéz:

Pocházejí dva vzorky z téhož základního souboru?

Nedopustili jsme se při dvou měřeních, jejichž výsledkem bylo určení dvou středních hodnot m_1, m_2, *systematických chyb?*

Má určitý faktor vliv na zkoumaný argument? Zde zkoumáme dva vzorky - jeden při působení daného faktoru, druhý bez jeho působení.

Řešené úlohy

Příklad 12.3.2. Odběratel dostává zářivky od dvou dodavatelů. Při hodnocení kvality zářivek se sleduje také počet zapojení, který snesou zářivky bez poškození. Zkoušky výrobků vedly k těmto výsledkům:

dodavatel *A*: 2139 2041 1968 1903 1952 1980 2089 1915

2389 2163 2072 1712 2018 1792 1849

dodavatel *B*: 1947 1602 1906 2031 2072

1812 1942 2074 2132

Ověřte hypotézu, že kvalita obou dodávek je stejná. Hladinu významnosti volte $p = 0,05$.

Řešení: V Excelu vypočteme charakteristiky obou souborů:

$n_1 = 15$ $M_1 = 1998,8$ $S_1^2 = 25444,69$

$n_2 = 9$ $M_2 = 1946,4$ $S_2^2 = 23554,25$

Nejdříve provedeme *F*-test:

Testovací kritérium:

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{n_1\ n_2-1\ .S_1^2}{n_2\ n_1-1\ .S_2^2} = \frac{15.\ 9-1\ .25444,69}{9.\ 15-1\ .23554,25} \doteq 1,0288$$

Kritická hodnota:

$F_{0,025}(14,8)$ = FINV(0,025;14;8) = 4,1297

Přijmeme tedy hypotézu o shodě rozptylů $\sigma_1^2 = \sigma_2^2$.

Dále tedy postupujeme jako v případě *a*):

Testovací kritérium:

$$T = \frac{M_1 - M_2}{\sqrt{n_1.S_1^2 + n_2.S_2^2}}.\sqrt{\frac{n_1.n_2.\ n_1 + n_2 - 2}{n_1 + n_2}} =$$

$$= \frac{1998,8 - 1946,4}{\sqrt{15.25444,69 + 9.23554,25}}.\sqrt{\frac{15.9.\ 15 + 9 - 2}{15 + 9}} \doteq 0,756$$

Kritická hodnota:

$t_{0,05}(22)$ = TINV(0,05;22) = 2,074

Závěr:

Testovací kritérium nepřekročilo kritickou hodnotu, přijmeme H_0: $\mu_1 = \mu_2$. Kvalita obou dodávek je stejná.

Tato úloha se dá v Excelu řešit i jednodušším způsobem, máme-li nainstalován doplňkový nástroj Excelu Analýza dat (instalace je podrobněji popsáno v 7.kapitole, příkladu 7.3.1.). Tento doplněk by mělo být možné spustit z nabídky Nástroje.

V dialogovém okně Analýza dat klepneme na analytický nástroj Dvouvýběrový *t*-test s rovností rozptylů. Objeví se nám okno, do kterého zadáme vstupy, tj. 1. soubor hodnoty od dodavatele *A*, 2. soubor hodnoty od dodavatele *B*. Výstupem pak bude následující (nebo velmi podobná) tabulka:

	dodav A	*dodav B*
Stř. hodnota	1998,8	1946,444
Rozptyl	27262,17	26498,53
Pozorování	15	9
Společný rozptyl	26984,48	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	22	
t stat	0,755905	
P(T<=t) (1)	0,228862	
t krit (1)	1,717144	
P(T<=t) (2)	0,457724	
t krit (2)	2,073875	

V této tabulce máme všechny potřebné údaje.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Příklad 12.3.3. Při antropologických měřeních obyvatelstva Egypta byla mimo jiné sledována šířka nosu (cm) u skupiny mužů 21-50 letých na severní části země a u skupiny stejně starých mužů z jižní části. Naměřené výsledky viz v tabulce. Posuďte významnost rozdílu ve výsledcích. Hladinu významnosti volte p = 0,05.

sever	3,6	4,1	3,3	3,4	3,7	3,1	4,0	4,0	3,6	3,0	3,3	
	3,7	4,3	3,3	3,4	3,4	3,3	3,6	4,0	3,4	3,7		
jih	4,3	3,9	4,3	3,8	4,1	4,2	3,8	3,9	3,8	3,8	4,0	3,7
	3,9	4,4	3,7	3,8	3,9	3,9	4,0	4,1	3,8	4,0	4,3	

Řešení: V Excelu vypočteme charakteristiky obou souborů:

$n_1 = 21$ $M_1 = 3{,}580952$ $S_1^2 = 0{,}112971$

$n_2 = 23$ $M_2 = 3{,}973913$ $S_2^2 = 0{,}0429249$

Nejdříve provedeme F-test:

Po dosazení do testovacího kritéria vyšla hodnota:

$F = 2{,}763409$

Kritická hodnota:

$F_{0,025}(20,22)$ = FINV(0,025;20;22) = 2,38898

Tudíž nemůžeme přijmout hypotézu o shodě rozptylů: $\sigma_1^2 \neq \sigma_2^2$.

Dále tedy postupujeme jako v případě *b*):

Testovací kritérium:

$$T = \frac{M_1 - M_2}{\sqrt{(n_2-1).S_1^2 + (n_1-1).S_2^2}}.\sqrt{(n_1-1).(n_2-1)} =$$

$$= \frac{3{,}580952 - 3{,}973913}{\sqrt{(23-1).0{,}112971 + (21-1).0{,}041059}}.\sqrt{(21-1).(23-1)} =$$

$$= -4{,}53304$$

Kritická hodnota, po dosazení:

$$t_p = \frac{(n_2-1).S_1^2.t_p(n_1-1) + (n_1-1).S_2^2.t_p(n_2-1)}{(n_2-1).S_1^2 + (n_1-1).S_2^2} \approx 2{,}083$$

Závěr:

Testovací kritérium v absolutní hodnotě překročilo kritickou hodnotu, nemůžeme přijmout H_0. Šířky nosu na severu se liší od těch na jihu.

Stejně jako u předchozí úlohy můžeme vyřešit v Excelu i pomocí doplňkového nástroje Analýza dat. V dialogovém okně Analýza dat klepneme na analytický nástroj Dvouvýběrový *t*-test s nerovností rozptylů. Objeví se nám okno, do kterého zadáme vstupy, tj. 1. soubor hodnoty ze severní části země, 2. soubor hodnoty z jihu.

Výstupem bude opět následující (nebo velmi podobná) tabulka:

	sever	*jih*
Stř. hodnot	3,580952	3,973913043
Rozptyl	0,118619	0,042924901
Pozorován	21	23
Hyp. rozdíl	0	
Rozdíl	32	
t stat	-4,53304	
P(T<=t) (1)	3,84E-05	
t krit (1)	1,693888	
P(T<=t) (2)	7,68E-05	
t krit (2)	2,036932	

V této tabulce opět najdeme všechny potřebné údaje.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

12.3.3. Studentův test pro párované hodnoty

Předpoklady:

Ze dvou normálně rozložených základních souborů s parametry μ_1, σ_1^2 a μ_2, σ_2^2 byly vybrány dva výběry se stejnými rozsahy n. Přitom každému prvku prvého výběru x_{1i} odpovídá právě jeden prvek druhého výběru x_{2i}. Vznikly tedy páry (x_{1i} ; x_{2i}), $i = 1, \dots n$.

Nulová hypotéza:

H_0: $\mu_1 = \mu_2$, což lze jinak zapsat: $\overline{d} = 0$, když $\overline{d}$ je střední hodnota rozdílů $d_i = x_{1i} - x_{2i}$, tedy:

$$\overline{d} = \frac{\sum_i x_{1i} - x_{2i}}{n} = \overline{x}_1 - \overline{x}_2 = 0\,.$$

Alternativní hypotéza:

H_1: $\mu_1 \neq \mu_2$ nebo tedy: $\overline{d} \neq 0$

Testovací kritérium:

$$t = \frac{\overline{d}.\sqrt{n-1}}{s_d}$$

(s_d je směrodatná odchylka hodnot d_i)

Veličina t má Studentovo rozložení s n - 1 stupni volnosti $t(n - 1)$.

Závěr:

Jestliže $|\, t \,| > t_p(n - 1)$, zamítneme hypotézu H_0.

Řešené úlohy

Příklad 12.3.4. Stanovení thiocyanového iontu (SCN-) bylo paralelně provedeno dvěma metodami (Aldridge a Barker) na 12 vzorcích. Srovnejte obě metodiky otestováním výsledků. Hladina významnosti p = 0,05.

	1	2	3	4	5	6	7	8	9	10	11	12
Aldridge	0,38	0,56	0,45	0,49	0,38	0,41	0,6	0,36	0,26	0,41	0,43	0,4
Barker	0,39	0,58	0,44	0,52	0,41	0,45	0,59	0,37	0,28	0,42	0,42	0,38

Řešení: Nejprve vytvoříme veličinu d:

Aldridge	0,38	0,56	0,45	0,49	0,38	0,41	0,6	0,36	0,26	0,41	0,43	0,4
Barker	0,39	0,58	0,44	0,52	0,41	0,45	0,59	0,37	0,28	0,42	0,42	0,38
d_i	-0,01	-0,02	0,01	-0,03	-0,03	-0,04	0,01	-0,01	-0,02	-0,01	0,01	0,02

Z tabulky jednoduše vypočteme potřebné charakteristiky:

$$\bar{d} = \frac{\sum_i d_i}{n} = \frac{-0,12}{12} = -0,01$$

(nebo v Excelu pomocí funkce PRŮMĚR)

Obdobně směrodatnou odchylku:

s_d = 0,018257

Testovací kritérium:

$$t = \frac{\left|\bar{d}\right|.\sqrt{n-1}}{s_d} = \frac{0,01.\sqrt{11}}{0,018257} \doteq 1,8166$$

Kritická hodnota:

$t_{0,05}(12 - 1)$ = TINV(0,05;11) = 2,201

Testovací kritérium nepřekročilo kritickou hodnotu, přijmeme H_0. Obě metodiky dávají stejné výsledky.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Přejděme nyní k ukázkám testů neparametrických, u nichž se nezaměřujeme na hodnoty některých parametrů základního souboru, ale studujeme shodu rozložení náhodné veličiny. Ověřujeme tedy např., zda určitý teoretický základní soubor může být modelem pro studovaný výběr, zda rozložení těchto souborů je možno považovat za totožná. Předveďme některé **testy dobré shody**.

12.4. Testy dobré shody (testy přiléhavosti)

12.4.1. Pearsonův test dobré shody - χ^2 test pro jeden výběr

Předpoklady:

Nechť výsledky pozorování jsou roztříděny do k skupin a v každé skupině je zjištěna skupinová četnost n_{ej} (četnosti experimentální). Uvažujme určité rozdělení, které budeme považovat za model pro náš výběr. Pro každou třídu určíme teoretické, modelové, očekávané četnosti n_{oj} (j = 1,...,k).

Nulová hypotéza:

H_0: Základní soubor má očekávané rozložení, tzn. že četnosti n_{ej} a n_{oj} (j = 1,...,k) se liší pouze náhodně.

Testovací kritérium:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(n_{ej} - n_{oj}\right)^2}{n_{oj}}$$

Tato veličina má Pearsonovo rozložení χ^2 s $\nu = k - s - 1$ stupni volnosti. Veličina s značí počet parametrů očekávaného rozložení odhadnutých na základě výběru.

Závěr:

Jestliže $\chi^2 > \chi_p^2(k - s - 1)$, zamítneme hypotézu H_0.

Poznámky

Při použití tohoto testu se vyžaduje splnění těchto podmínek:

- *všechny očekávané třídní četnosti mají být větší než 1,*
- *nejvýš 20 % očekávaných třídních může být menších než 5,*
- *nedoporučuje se volit počet tříd větší než 20.*

Nejsou-li splněny, lze přikročit k sloučení sousedních tříd v nezbytném rozsahu. Pozn. ke stupňům volnosti: Ověřujeme-li např. normalitu základního souboru, je **s** rovno 2, *protože teoretické normální rozložení se stanovuje na základě odhadu střední hodnoty a disperze výběru, tedy na základě dvou charakteristik.*

Řešené úlohy

Příklad 12.4.1. Je dán statistický soubor. Na hladině významnosti 5 % otestujte hypotézu, že soubor má normální rozdělení.

i	1	2	3	4	5	6	7	8	9	10	11	12
obsah Al_2O_3	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20
n_{ei}	2	5	7	19	52	57	72	61	19	14	4	1

Řešení: Nejdříve vypočteme příslušné charakteristiky, tj. parametry normálního rozdělení - střední hodnotu a rozptyl. Výpočet provedeme způsobem, který byl popsán v 7. kapitole, příkladu 7.4.1.:

i	obsah Al_2O_3	třídní znak x_i	f_i	$x_i.f_i$	$(x_i-M)^2.f_i$
1	8 - 9	8,5	2	17	63,02094
2	9 - 10	9,5	5	47,5	106,4182
3	10 - 11	10,5	7	73,5	91,39755
4	11 - 12	11,5	19	218,5	129,7692
5	12 - 13	12,5	52	650	135,3622
6	13 - 14	13,5	57	769,5	21,44809
7	14 - 15	14,5	72	1044	10,76006
8	15 - 16	15,5	61	945,5	117,2791
9	16 - 17	16,5	19	313,5	108,2197
10	17 - 18	17,5	14	245	160,5651
11	18 - 19	18,5	4	74	76,96839
12	19 - 20	19,5	1	19,5	29,01526
Σ		***N*** **=**	**313**	**4417,5**	**1050,224**
				14,11342	3,272014
					1,808871

Střední hodnota:

$$M = \frac{1}{N}\sum_i x_i f_i = \frac{4417{,}5}{313} = 14{,}11342$$

Rozptyl:

$$S^2 = n_2 = n_2 - \frac{h^2}{12} = \frac{1}{N}\sum_i (x_i - M)^2 f_i - \frac{h^2}{12} =$$

$$= \frac{1050,224}{313} - \frac{1}{12} = 3,272014$$

Směrodatná odchylka:

$$S = \sqrt{3,272014} = 1,808871$$

Pomocí parametrů normálního rozdělení můžeme vypočítat očekávané četnosti n_{oi}:
Uvedeme např. výpočet n_{o1}:

$n_{o1} = N.P(8 \le X \le 9) = 313.(F(9) - F(8))$ = (v Excelu) =

= 313*(NORMDIST(9;14,11342;1,808871;1) -

- NORMDIST(8;14,11342;1,808871;1)) =

= 0,6220961

Zbylé očekávané četnosti vypočteme analogicky, viz. tabulka:

i	obsah Al_2O_3	n_{ei}	n_{oi}
1	8 - 9	2	0,6220961
2	9 - 10	5	2,8582712
3	10 - 11	7	9,7422953
4	11 - 12	19	24,64009
5	12 - 13	52	46,25248
6	13 - 14	57	64,446882
7	14 - 15	72	66,661732
8	15 - 16	61	51,187338
9	16 - 17	19	29,176478
10	17 - 18	14	12,343305
11	18 - 19	4	3,8750334
12	19 - 20	1	0,9025231

Z tabulky je patrné, že nejsou splněny všechny podmínky z předchozí poznámky, proto sloučíme třídy 1,2 a třídy 11,12:

i	obsah Al_2O_3	n_{ei}	n_{oi}	n_{ei} po sloučení tříd	n_{oi}	$(n_{ei}-n_{oi})^2/n_{oi}$
1	8 - 9	2	0,6220961	7	3,480367	3,559341319
2	9 - 10	5	2,8582712			
3	10 - 11	7	9,7422953	7	9,742295	0,771910871
4	11 - 12	19	24,64009	19	24,64009	1,291010614
5	12 - 13	52	46,25248	52	46,25248	0,714209942
6	13 - 14	57	64,446882	57	64,44688	0,860492344
7	14 - 15	72	66,661732	72	66,66173	0,427488209
8	15 - 16	61	51,187338	61	51,18734	1,881096963
9	16 - 17	19	29,176478	19	29,17648	3,549458607
10	17 - 18	14	12,343305	14	12,34331	0,222358324
11	18 - 19	4	3,8750334	5	4,777557	0,010356987
12	19 - 20	1	0,9025231			
					TK:	13,28772418

Po sloučení tříd jsou všechny podmínky splněny, v posledním sloupci je vypočtena hodnota testovacího kritéria:

$$\chi^2 = \sum_i \frac{\left(n_{ei} - n_{oi}\right)^2}{n_{oi}} = 13,2877$$

Kritická hodnota:

$$\chi^2_{0,05}\left(10-2-1\right) = \chi^2_{0,05}\left(7\right) = \text{CHIINV(0,05;7)} = 14,067$$

Závěr:

Testovací kritérium nepřekročilo kritickou hodnotu. Daný soubor má normální rozdělení.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

12.4.2. Kolmogorovův-Smirnovův test dobré shody pro jeden výběr

Předpoklady:

Nechť výsledky pozorování jsou roztříděny do k skupin a v každé skupině je zjištěna skupinová četnost n_{ej} (četnosti experimentální). Uvažujme určité rozdělení, které budeme považovat za model pro náš výběr. Pro každou třídu určíme teoretické, modelové, očekávané četnosti n_{oj} ($j = 1,...,k$).

Pro empirické i teoretické očekávané rozdělení stanovíme kumulativní četnosti N_{ej} a N_{oj}, $j = 1,...,k$.

Nulová hypotéza:

H_0: Základní soubor má očekávané rozložení, tzn. že četnosti N_{ej} a N_{oj} ($j = 1,...,k$) se liší pouze náhodně.

Testovací kritérium:

$$D_1 = \frac{1}{n}.\max\left|N_{ej} - N_{oj}\right|, \quad j = 1, \cdots, k$$

Tato veličina má speciální rozložení, jehož kritické hodnoty jsou tabelovány pro $n < 40$ (viz tabulky). Pro $n \geq 40$ se počítají podle přibližných vzorců.

Pro hladinu významnosti p = 0,05 je

$$D_{1;0,05}\ n\ = \frac{1,36}{\sqrt{n}},$$

pro hladinu významnosti p = 0,01 je

$$D_{1;0,01}\ n\ = \frac{1,63}{\sqrt{n}}.$$

Závěr:

Jestliže $D_1 \geq D_{1;p}$, zamítneme hypotézu H_0.

Řešené úlohy

Příklad 12.4.2. Využijeme zadání příkladu 12.4.1. a úlohu vyřešíme pomocí Kolmogorovova - Smirnovova testu pro jeden výběr:

Řešení: Parametry normálního rozdělení a očekávané četnosti jsme už vypočetli v příkladě 12.4.1., stačí dopočítat kumulativní četnosti a testovací kritérium:

i	obsah Al_2O_3	n_{ei}	n_{ei} po sloučení	n_{oi}	N_{ei}	N_{oi}	$N_{ei} - N_{oi}$
1	8 - 9	2	7	3,480367	7	3,480367	3,519633
2	9 - 10	5					
3	10 - 11	7	7	9,742295	14	13,22266	0,777337
4	11 - 12	19	19	24,64009	33	37,86275	-4,86275
5	12 - 13	52	52	46,25248	85	84,11523	0,884767
6	13 - 14	57	57	64,44688	142	148,5621	-6,56212
7	14 - 15	72	72	66,66173	214	215,2238	-1,22385
8	15 - 16	61	61	51,18734	275	266,4112	**8,588815**
9	16 - 17	19	19	29,17648	294	295,5877	-1,58766
10	17 - 18	14	14	12,34331	308	307,931	0,069032
11	18 - 19	4	5	4,777557	313	312,7085	0,291476
12	19 - 20	1					
		n =	313			TK:	**0,02744**

Testovací kritérium:

$$D_1 = \frac{1}{n}.\max\left|N_{ei} - N_{oi}\right| = \frac{8,588815}{313} = 0,02744 .$$

Kritická hodnota:

$$D_{1;0,05}\ \ 313\ = \frac{1,36}{\sqrt{313}} = 0,076872 .$$

Testovací kritérium nepřekročilo kritickou hodnotu. Daný soubor má normální rozdělení.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Předchozí dva testy ověřovaly, zda rozložení výběru neodporuje předpokladu o určitém rozložení základního souboru. Následující test bude ověřovat, shodu rozložení dvou výběrů.

12.4.3. Kolmogorovův-Smirnovův test dobré shody pro dva výběry

Předpoklady:

U dvou výběrových souborů s rozsahy n_1 a n_2 bylo provedeno roztřídění do k skupin a zjištěny kumulativní třídní četnosti pro každou třídu: $N_{1,j}$ a $N_{2,j}$. $F_{1,j}$ a $F_{2,j}$ jsou pak příslušné třídní relativní kumulativní četnosti.

Nulová hypotéza:

Oba výběrové soubory mají totéž rozložení (pocházejí tedy z téhož základního souboru).

Testovací kritérium:

a) $n_1 = n_2 \leq 40$

$$D_2 = \max_j \left|N_{1j} - N_{2j}\right|,\ \ j = 1,\cdots,k$$

má speciální rozložení, jeho kritické hodnoty se vyčtou z příslušných tabulek (viz tabulky),

b) $n_1 > 40$ a $n_2 > 40$ (i různě velké):

$$D_2 = \max_j \left|F_{1j} - F_{2j}\right|,\ \ j = 1,\cdots,k .$$

Kritické hodnoty se počítají podle vzorců:

pro $p = 0,05$ je

$$D_{2;0,05} = 1,36.\sqrt{\frac{n_1 + n_2}{n_1.n_2}}\ \text{a}$$

pro p = 0,01 je

$$D_{2;0,01} = 1,63.\sqrt{\frac{n_1 + n_2}{n_1.n_2}}\ .$$

Závěr:

Jestliže $D_2 \geq D_{2;p}(n_1,n_2)$, zamítneme nulovou hypotézu H_0.

Řešené úlohy

Příklad 12.4.3. Ve dvaceti vybraných závodech byly zkoušeny dva typy filtrů odpadních vod. Bylo zjišťováno, jaké procento nečistot filtr zadrží, a to tak, že nejprve byly instalovány filtry 1. typu a po určité době filtry 2. typu. Výsledky jsou v tabulce. Zjistěte, jestli se porovnávané filtry kvalitativně liší.

množství zadržených nečistot (v %)	10	20	30	40	50	60	70
$n_{1,j}$	1	2	3	8	5	1	0
$n_{2,j}$	0	2	3	2	3	7	3

Řešení:

H_0: Dva základní soubory mají totéž rozdělení (porovnávané filtry se kvalitativně neliší).

Volíme hladinu významnosti p = 0,05

množství zadržených nečistot (v %)	$n_{1,j}$	$n_{2,j}$	$N_{1,j}$	$N_{2,j}$	$\lvert N_{1,j} - N_{2,j}\rvert$
10	1	0	1	0	1
20	2	2	3	2	1
30	3	3	6	5	1
40	8	2	14	7	7

50	5	3	19	10	9
60	1	7	20	17	3
70	0	3	20	20	0
$\Sigma =$	20	20			

Z tabulky vidíme, že $n_1 = n_2 < 40$, tudíž testovací kritérium:

$$D_2 = \max_j \left| N_{1,j} - N_{2,j} \right| = 9$$

Kritická hodnota:

$D_{2;0,05}(20) = 9$ (viz tabulky)

Závěr:

$D_2 = D_{2;0,05}(20) = 9$, zamítneme H_0.

Filtry se kvalitativně liší.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Existují i neparametrické testy, které neověřují rozložení výběrového souboru. Uveďme test, který se snaží zjistit, zda výběrový soubor neobsahuje údaj zatížený hrubou chybou měření, popř. chybou v zápise. Jde o jeden z **testů extrémních odchylek**.

12.5. Testy extrémních hodnot

12.5.1. Dixonův test extrémních odchylek

Předpoklady:

Ve výběrovém souboru o rozsahu n je $x_1 = \min(x_i)$, resp. $x_n = \max(x_i)$ (např. hodnoty jsou seřazeny podle velikosti od x_1 do x_n).

Nulová hypotéza:

H_0: Hodnota x_1 (nejmenší hodnota), resp. x_n (největší hodnota) se neliší významně od ostatních hodnot souboru.

Testovací kritérium:

$$Q_1 = \frac{x_2 - x_1}{x_n - x_1}, \text{ nebo } Q_n = \frac{x_n - x_{n-1}}{x_n - x_1},$$

podle toho, testujeme-li minimální nebo maximální hodnotu ve výběru. Kritické hodnoty $Q_{1;p}$, resp. $Q_{n;p}$ se vyčtou z příslušných tabulek (viz tabulky).

Závěr:

Jestliže $Q_1 > Q_{1;p}$, resp. $Q_n > Q_{n;p}$, zamítneme nulovou hypotézu H_0.

Test extrémních odchylek je možno ovšem také provést užitím parametrického testu:

12.5.2. Grubbsův test extrémních odchylek

Předpoklady:

Ve výběrovém souboru o rozsahu n je $x_1 = \min(x_i)$, resp. $x_n = \max(x_i)$ (např. hodnoty jsou seřazeny podle velikosti od x_1 do x_n). $\bar{x}$ je střední hodnota výběru, S je výběrová směrodatná odchylka.

Nulová hypotéza:

H_0: Hodnota x_1, resp. x_n se neliší významně od ostatních hodnot souboru.

Testovací kritérium:

$$T_1 = \frac{\bar{x} - x_1}{S}, \text{ resp. } T_n = \frac{x_n - \bar{x}}{S},$$

podle toho, testujeme-li minimální nebo maximální hodnotu ve výběru.

Kritické hodnoty $T_{1;p}$, resp. $T_{n;p}$ se vyčtou z příslušných tabulek (viz tabulky),

Závěr:

Jestliže $T_1 > T_{1;p}$, resp. $T_n > T_{n;p}$, zamítneme nulovou hypotézu H_0.

Poznámka

Vede-li test k závěru, že extrémní hodnotu je třeba ze souboru vyloučit, je třeba sestrojit znovu všechny výběrové charakteristiky (ze souboru bez extrémní hodnoty) pro případné další výpočty.

Řešené úlohy

Příklad 12.5.1. Při kalibraci titrační metody k stanovení krevního cukru bylo provedeno 12 paralelních analýz z jednoho vzorku s výsledky v tabulce. Otestujte, zda hodnota 98 není chybná.

83	88	84	78	82	82
86	81	98	83	85	80

Dixonovým testem:

x_1 = 78 (nejmenší hodnota)

x_{n-1} = 88 (druhá největší hodnota)

Testovací kritérium:

$$Q_n = \frac{x_n - x_{n-1}}{x_n - x_1} = \frac{98-88}{98-78} = 0,5$$

Kritická hodnota:

$Q_{12;0,05} = 0,376$;

$Q_{12;0,01} = 0,482$ (viz tabulky).

Závěr:

Testovací kritérium překročilo kritickou hodnotu (pro obě zkoumané hladiny významnosti). Zamítáme nulovou hypotézu H_0.

Hodnota 98 se významně liší od ostatních hodnot.

Grubbsovým testem:

Nejdříve vypočteme potřebné charakteristiky:

$\bar{x} = 84,16667$ $S = 4,896144$

Testovací kritérium:

$$T_n = \frac{x_n - \bar{x}}{S} = \frac{98-84,16667}{4,896144} \doteq 2,825$$

Kritická hodnota:

$Q_{12;0,05} = 2,387$;

$Q_{12;0,01} = 2,663$ (viz tabulky).

Závěr:

Testovací kritérium překročilo kritickou hodnotu (pro obě zkoumané hladiny významnosti). Zamítáme nulovou hypotézu H_0.

Hodnota 98 se významně liší od ostatních hodnot.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

Uveďme ještě test, který se týká koeficientu korelace u dvojrozměrné náhodné veličiny.

12.6. Testy o koeficientu korelace

12.6.1. Test lineární nezávislosti v základním souboru

Předpoklady:

Dvojrozměrný základní soubor má normální rozložení a korelační koeficient $\boldsymbol{\rho}$.

Náhodný výběr z tohoto souboru má rozsah $\boldsymbol{n}$ a koeficient korelace $\boldsymbol{R}$.

Nulová hypotéza:

$\rho = 0$

Testovací kritérium:

$$t = \frac{R}{\sqrt{1-R^2}}.\sqrt{n-2}$$

Tato veličina má Studentovo rozložení s n - 2 stupni volnosti $t(n - 2)$.

Závěr:

Jestliže $|t| > t_p\ \ n-2$, zamítneme H_0.

Poznámka

Odmítnutí nulové hypotézy znamená připuštění alternativní hypotézy, že mezi složkami náhodné veličiny je korelace, nejsou lineárně nezávislé.

Řešené úlohy

Příklad 12.6.1. Otestujte na hladině významnosti $p = 0{,}05$, zda u dvojrozměrné veličiny dané v tabulce, může jít o lineární závislost.

x	0,0	0,5	1,0	1,5	2,0	2,5	3,0
y	0,0	1,7	3,1	3,8	3,9	3,8	3,0

Řešení: Použijeme předchozí test lineární nezávislosti v základním souboru.

Nejdříve (např. v Excelu vypočteme výběrový koeficient korelace:

$R = 0{,}752064$.

Tuto hodnotu dosadíme do testovacího kritéria:

$$t = \frac{R}{\sqrt{1-R^2}}.\sqrt{n-2} = \frac{0{,}752064}{\sqrt{1-0{,}752064^2}}.\sqrt{7-2} \doteq 2{,}551495.$$

Kritická hodnota:

$t_{0,05}(7\text{-}2)$ = TINV(0,05;D22) = 2,570582.

Závěr:

Hodnota testovacího kritéria nepřekročila kritickou hodnotu.

Není nutno zamítnout hypotézu o lineární nezávislosti x a y.

Tuto úlohu si můžete otevřít **vyřešenou** v Excelu.

K procvičení předchozích poznatků si otevřete **sbírku úloh**, ve které najdete mnoho řešených i neřešených příkladů z matematické statistiky.

Úlohy k samostatnému řešení

12.1. Dva automaty vyrábějí součástky téhož druhu. Ze součástek vyrobených na prvním automatu jsme změřili $n_1 = 9$ součástek, ze součástek vyrobených na druhém automatu $n_2 = 12$ součástek. Výběrové disperze měřené délky jsou $s_1^2 = 6$ μm, $s_2^2 = 23$ μm. Můžeme přijmout hypotézu o rovnosti disperzí na hladině významnosti 0,05?

12.2. Každé ze dvou polí bylo rozděleno na 10 lánů a zaseto obilí. Přitom na lánech prvního pole bylo použito speciální americké hnojivo. Výnosy z lánů prvního a druhého pole měly průměry $\overline{x_1} = 6$; $\overline{x_2} = 5{,}7$ a rozptyly $s_1^2 = 0{,}064$; $s_2^2 = 0{,}024$. Zjistěte na 5% hladině významnosti, jestli hnojení mělo průkazný vliv na výnosy.

12.3. Dvě skupiny studentů prováděly shyby na hrazdě s těmito výsledky:

I. skupina:

počet shybů	0	3	5	6	7	8	9	10
četnost	2	2	3	8	7	4	3	1

II. skupina:

počet shybů	4	5	6	7	8	9	10
četnost	1	4	5	8	8	2	2

Proveďte F-test pro $p = 0{,}05$.

12.4. U dvou vzorků byly změřeny základní charakteristiky: $n_1 = 10$, $\overline{x_1} = 26{,}5$; $s_1^2 = 4{,}5$; $n_2 = 5$, $\overline{x_2} = 28$; $s_2^2 = 5{,}8$. Jsou střední hodnoty obou vzorků významně odlišné na hladině významnosti 5 %?

12.5. U dvou vzorků byly změřeny základní charakteristiky: $n_1 = 10$, $\overline{x_1} = 18$; $s_1^2 = 0{,}85$; $n_2 = 6$, $\overline{x_2} = 14$; $s_2^2 = 0{,}22$. Jsou střední hodnoty obou vzorků významně odlišné na hladině významnosti 5 %?

12.6. Svaly horní končetiny byly cyklicky namáhány až do úplného vypovězení funkce. Hmotnost závaží byla konstantní a délka přestávky mezi sériemi byla 30 sekund. Otestujte, zda jsou obě končetiny stejně silné.

série	1	2	3	4	5	6	7	8	9	10	11
končetina P	20	7	3	2	2	2	1	1	1	0	0
končetina L	19	6	3	3	2	2	2	1	1	1	0

12.7. Prověřte na 5% hladině významnosti, zda soubor má rovnoměrné rozdělení, když pro náhodný výběr byly zjištěny tyto četnosti jednotlivých tříd:
10, 21, 0, 8, 12, 6, 8, 13, 11, 11.

12.8. Zjistěte, zda nejmenší hodnota v daném souboru je extrémně odchýlena od ostatních. Hladinu významnosti volte $p = 0{,}05$. Testovaný soubor:

111,2	112,4	114,6	95,4	105,6	107,7	108,3	111,8	115,3	109,1

Výsledky úloh k samostatnému řešení

12.1. ano

12.2. ano

12.3. zamítáme nulovou hypotézu

12.4. ne

12.5. ano

12.6. obě končetiny jsou stejně silné

12.7. nemá

12.8. je extrémně odchýlená

PRAVDĚPODOBNOST A STATISTIKA - SBÍRKA ÚLOH

Úlohy k samostatnému řešení

(Odkazy ukazují na sešity excelu, v nichž jsou uvedené příklady vyřešeny, pokud není uvedeno, že jde o "zadání". V jednom sešitě může být uvedeno více příkladů. Text příkladů je možno zkopírovat do vlastního sešitu excelu a řešit úlohy samostatně. Některé příklady byly uvedeny v předešlém textu.)

(0020.xls)

Byly sledovány výsledky běhu na 50 m (ve vteřinách) u skupiny desetiletých chlapců a dívek. Posuďte získané výsledky z hlediska vyrovnanosti výkonů v jednotlivých skupinách.

Chlapci:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
10,80	9,30	9,40	9,90	10,20	9,30	9,40	8,90	8,90	9,60	9,70	10,60	9,40	9,50	9,60	10,00	9,30

18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
9,40	8,40	9,80	8,80	9,20	9,50	9,80	9,00	10,50	9,40	9,30	9,90	9,10	9,60	8,70	8,10

Dívky:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
10,70	10,80	10,00	10,60	9,20	10,20	9,90	10,00	9,30	10,20	9,80	10,00	10,00	11,00

15	16	17	18	19	20	21	22	23	24	25	26	27	28
12,00	10,00	10,00	11,20	9,40	10,70	9,30	10,10	9,10	10,20	9,30	10,00	9,40	10,90

(0021.xls)

Odběratel dostává zářivky od dvou dodavatelů. Při hodnocení kvality zářivek se sleduje také počet zapojení, která snesou zářivky bez poškození. Zkoušky výrobků vedly k těmto výsledkům:

dodavatel *A*: 2139 2041 1968 1903 1952 1980 2089 1915

2389 2163 2072 1712 2018 1792 1849

dodavatel *B*: 1947 1602 1906 2031 2072

1812 1942 2074 2132

Ověřte hypotézu, že kvalita obou dodávek je stejná. Hladinu významnosti volte $p = 0{,}05$.

(0022.xls)

Při antropologických měřeních obyvatelstva Egypta byla mimo jiné sledována šířka nosu (cm) u skupiny mužů 21-50 letých na severní části země a u skupiny stejně starých mužů z jižní části. Naměřené výsledky viz v tabulce. Posuďte významnost rozdílu ve výsledcích. Hladinu významnosti volte $p = 0{,}05$.

sever 3,6 4,1 3,3 3,4 3,7 3,1 4,0 4,0 3,6 3,0 3,3

3,7 4,3 3,3 3,4 3,4 3,3 3,6 4,0 3,4 3,7

jih 4,3 3,9 4,3 3,8 4,1 4,2 3,8 3,9 3,8 3,8 4,0 3,7

3,9 4,4 3,7 3,8 3,9 3,9 4,0 4,1 3,8 4,0 4,3

(0023.xls)

Stanovení thiocyanového iontu (SCN-) bylo paralelně provedeno dvěma metodami (Aldridge a Barker) na 12 vzorcích. Srovnejte obě metodiky otestováním výsledků. Hladina významnosti $p = 0{,}05$.

	1	2	3	4	5	6	7	8	9	10	11	12
Aldridge	0,38	0,56	0,45	0,49	0,38	0,41	0,6	0,36	0,26	0,41	0,43	0,4
Barker	0,39	0,58	0,44	0,52	0,41	0,45	0,59	0,37	0,28	0,42	0,42	0,38

(0025.xls)

Při sériové výrobě určitého předmětu byly na podkladě kontrolních měření zjišťovány vadné výrobky vyrobené v každé hodině během jedné směny. Ověřte, zda výskyt vadných výrobků během směny je rovnoměrný.

hodina výroby	1	2	3	4	5	6	7	8
počet zmetků	29	7	27	61	87	110	101	42

(0026.xls)

Otestujte na hladině významnosti p = 0,05 hypotézu, že základní soubor, z něhož jsme vybrali vzorek, má normální rozložení. Variační řada je dána tabulkou:

x	220	230	240	250	260	270	280
f_x	2	5	25	38	20	7	3

(0027.xls)

Najděte korelační matici pro dvojrozměrný statistický soubor daný četnostní tabulkou:

$x \setminus y$	20	30	40	50	60	70	80
250	19	5					
350	23	116	11				
450	1	41	98	9			
550		4	32	65	7		
650		1	4	21	46	3	
750			1	2	11	13	1
850					1	3	2

(0028.xls)

Určete oboustranný konfidenční interval rozptylu normálně rozloženého základního souboru pro hladiny spolehlivosti 0,90; 0,95 a 0,99, když u výběru s rozsahem $n = 12$ byl zjištěn rozptyl 0,64. Posuďte získané výsledky.

(0029.xls)

Měřili jsme průměr vačkového hřídele na 250 součástkách. Předpokládáme normální rozdělení souboru. Z výsledků měření jsme určili výběrový průměr a výběrovou disperzi $x_p = 995{,}6$, $s^2 = 134{,}7$. Určete interval spolehlivosti pro střední hodnotu základního souboru při hladině významnosti 5%.

(0029.xls)

Při měření kapacity sady kondenzátorů bylo provedeno 10 měření s výsledky:

152	156	148	153	150	156	140	155	145	148

Odhadněte interval spolehlivosti pro kapacitu těchto kondenzátorů se spolehlivostí 90 %, resp. 95 %.

(0029.xls)

Bylo zkoušeno 30 náhodně vybraných ocelových tyčí k určení meze kluzu určitého druhu oceli. Po zpracování výsledků byla určena její empirická střední hodnota 286,4 MPa a rozptyl 121 [MPa^2].

a) Určete intervalový odhad parametrů základního souboru s 95% spolehlivostí.

b) Kolik vzorků by bylo třeba zvolit, aby chyba určené střední hodnoty nepřesáhla 2 MPa?

(0031.xls)

Zpracování dvojrozměrného souboru daného lineární tabulkou hodnot.

x	27	31	87	93	114	124	190	193	250	254	264	272
y	28	21	71	36	30	43	54	54	59	25	82	22

308	324	371	372	440	442	502	503	506	522	556	620	624
38	22	56	63	46	24	33	40	41	28	53	38	66

(0030.xls)

Zpracování dvojrozměrného statistického souboru daného četnostní tabulkou.

$x \setminus y$	20	30	40	50	60	70	80
250	19	5					
350	23	116	11				
450	1	41	98	9			
550		4	32	65	7		
650		1	4	21	46	3	
750			1	2	11	13	1
850					1	3	2

(zadání 0033.xls)

Určete decily, kvantily a medián statistického souboru daného variační řadou:

a)

x_k	1	2	3	4	5	6	7
f_k	2	15	16	17	14	13	2

b)

x_k	2	3	4	5	6
f_k	6	11	18	12	8

(zadání 0033.xls)

Určete průměrnou dobu, kterou potřebuje k splnění úkolu družstvo vojáků, když vojáci *A* a *B* k tomu potřebovali 3 min., vojáci *C*, *D* 5 min. a voják *E* 6 min.

(zadání 0033.xls)

Řidič nákladního automobilu ujel 150 km, z toho 20 km rychlostí 30 km//h, 30 km rychlostí 40 km/h, 50 km rychlostí 60 km/h 10 km rychlostí 70 km/h. Určete průměrnou rychlost auta.

(zadání 0033.xls)

Určete variační interval, variační rozpětí, aritmetický průměr, rozptyl, směrodatnou odchylku a variační koeficient množství srážek naměřených (v mm) v Brně v období let 1941 až 1960.

718,5	492,3	431,5	540,5	514,7	584,0	385,0	532,0	531,0	578,3
551,9	613,6	476,0	661,3	518,0	508,5	488,7	494,9	554,6	673,5

(zadání 0033.xls)

Určete roční průměr, směrodatnou odchylku a variační koeficient průtoku Labe v r. 1968 na určitém místě, jsou-li známy měsíční průtoky (v m^3/sec):

40,7	57,9	121,0	74,8	51,6	45,5	41,4	87,7	56,8	129,0	99,2	125,0

(zadání 0033.xls)

Mnohonásobným měření byla zjištěna následující variační řada velikostí zatížení silničního mostu (v kp/m^2):

zatížení	300	350	400	450	500	550	600	650	700	750	800	Σ
f_k / n %	0	3,44	17,05	30,12	25,3	15,8	6,35	1,72	0,21	0,01	0	100

Vypočtěte statistické charakteristiky sledované veličiny.

(zadání 0033.xls)

Při prověrkách tělesné zdatnosti 100 branců se výkony ve skoku do dálky pohybovaly v rozmezí 380 až 580 cm. Výsledky jsou shrnuty v tabulce:

středy tříd	390	410	430	450	470	490	510	530	550	570
f_k	7	10	14	22	25	12	3	3	2	2

Určete všechny momentové charakteristiky tohoto souboru (příp. i s použitím Shepardových korekcí).

(0034.xls)

Při kalibraci titrační metody k stanovení krevního cukru bylo provedeno 12 paralelních analýz z jednoho vzorku s těmito výsledky:

83	88	84	78	82	82	86	81	98	83	85	80	(mg %)

Otestujte, zda hodnota 98 není chybná.

Nevěrohodnost minimálního obsahu byla zjištěna v souboru 10 silikátových analýz žul. Analýzou byly zjištěny následující obsahy SiO_2:

číslo vzorku	1	2	3	4	5	6	7	8	9	10
obsah SiO_2 v %	72,5	59,4	75,6	68,0	63,0	70,1	72,9	68,5	54,5	78,0

Můžeme výsledek 9. pozorování považovat za odlehlý?

(0036.xls)

Sledujte počty absolventů Zemědělské vysoké školy ve Vídni (University fur Bodenkultur) od školního roku 1929/30 do 1990/91 pro obor zemědělství.

42	56	36	46	45	35	50	46	39	31	49	5	10	17	20
36	65	74	144	129	128	88	63	72	51	42	58	47	35	28
41	34	50	57	54	48	61	45	53	47	31	50	53	25	41
34	39	51	36	45	34	67	89	78	77	116	81	98	90	145
110														

(0037.xls)

Určete elementární charakteristiky růstu časové řady sledující výrobu plynu v letech 1980 - 1985:

rok	1980	1981	1982	1983	1984	1985
výroba (m^3)	1286	1363	1393	1495	1571	1610

Náhodným výběrem o rozsahu n = 10 byly vybrány vzorky paliva o výhřevnosti (údaje v kJ/kg):

12 016	11 824	13 253	11 489	12 335	12 791	12 167	13 183	13 428	12 446

Ověřte na hladině významnosti 5 %, že uvedený výběr pochází ze základního souboru normálně rozloženého se střední hodnotou 12500 kJ/kg a směrodatnou odchylkou 1000 kJ/kg.

(zadání 0041.xls)

Byly vytvořeny dva soubory náhodných výběrů vzorků paliva o rozsahu $n_1 = n_2 =$ 100. U 1. vzorku byl zjištěn průměr 12 424 kJ/kg a směrodatná odchylka 902 kJ/kg. U 2. výběru průměr 12 526 kJ/kg a směrodatná odchylka 939 kJ/kg.

Rozhodněte na 5% hladině významnosti, zda tyto oba výběry pocházejí ze základního souboru se stejnou střední hodnotou.

(Přeformulujte úlohu více do jazyka technika než statistika, aby byl patrnější důvod provádění testu.)

(zadání 0041.xls)

Každé ze dvou polí bylo rozděleno na 10 lánů a zaseto obilí. Přitom na lánech prvního pole bylo použito speciální americké hnojivo. Výnosy z lánů prvního a druhého pole měly průměry $x_1 = 6$; $x_2 = 5{,}7$ a rozptyly $s_1^2 = 0{,}064$; $s_2^2 = 0{,}024$. Zjistěte na 5% hladině významnosti, jestli hnojení mělo průkazný vliv na výnosy.

(zadání 0041.xls)

Dva druhy ocelových pružin byly vyšetřovány z hlediska pevnosti v tahu. Bylo vyšetřeno n_1 = 145 pružin typu A a $n_2 = 200$ pružin typu B s těmito výsledky:

$m_1 = 31{,}40$ kp/mm^2, $s_1 = 3{,}26$ kp/mm^2, $m_2 = 29{,}84$ kp/mm^2, $s_2 = 3{,}51$ kp/mm^2.

Zjistěte, zda rozdílnost hodnot je náhodně vysvětlitelná.

(zadání 0041.xls)

Měřením téže veličiny dvěma přístroji A a B jsme během 8 dnů dostali u přístroje A hodnoty u_k a u přístroje B hodnoty v_k.

den k	1	2	3	4	5	6	7	8
u_k	51,8	54,9	52,2	53,3	51,6	54,1	54,2	53,3
v_k	49,5	53,3	50,6	52,0	46,8	50,5	52,1	53,0

Zjistěte, zda tyto hodnoty opravňují k domněnce, že kvality obou přístrojů se významně neliší.

(zadání 0041.xls)

Z výroby automatu vyrábějícího určité zboží byly vzaty v různých dobách dva vzorky o rozsahu $n_1 = n_2 = 5$, s průměry $m_1 = 20{,}096$, $m_2 = 20{,}084$, rozptyly $s_1^2 = 0{,}0013$, $s_2^2 = 0{,}0004$. Zjistěte, zda během uvedené doby zůstal automat stejně seřízen.

(zadání 0041.xls)

Jsou dány výsledky měření 1000 součástek se zaokrouhlením na 0,5 mm četnostní tabulkou:

i	1	2	3	4	5	6	7	8	9	10
x_i	98	98,5	99	99,5	100	100,5	101	101,5	102	102,5
f_i	21	47	87	158	181	201	142	97	41	25

Ověřte, zda získaná pozorování jsou v souhlase s předpokladem, že měřená veličina má normální rozložení.

(zadání 0041.xls)

Při 30 hodech hrací kostkou padla šestka čtyřikrát, při dalších 40 hodech sedmkrát. Rozhodněte na 1% hladině významnosti, zda je rozdíl v počtu padnuvších šestek statistický významný.

(zadání 0041.xls)

Zjistěte, zda hrací kostka je správná, zda tedy dává všem číslům stejnou naději, na základě 300 hodů s těmito výsledky:

x_i	1	2	3	4	5	6
f_i	64	55	41	53	40	47

(zadání 0041.xls)

Z 10 úseků rudného dolu bylo pro zjištění průměrné kovnatosti těžených hornin odebráno po jednom vzorku o váze 1t.

úsek	1	2	3	4	5	6	7	8	9	10
kovnatost	0,6	2,4	2,1	1,4	1,2	4,8	0,9	1,1	3,5	3,0

Ověřte hypotézu, že těžená kovnatost se neliší významně od plánované kovnatosti 2,7%

(zadání 0041.xls)

Při výpočtu zásob u Sn-rudy byly zjištěny škodlivé příměsi W, S, Bi, As. Obsah těchto příměsí je bedlivě sledován, neboť jejich zvýšený obsah nad přípustnou hranici má vliv na náklady upravárenského a hutnického procesu a tím na cenu ložiska.

U 10 analyzovaných vzorků vykázal jeden vzorek hodnotu 0,9 nad přípustnou mez 0,5 %. Ověřte, zda je nutno tuto hodnotu vyloučit.

vzorek	1	2	3	4	5	6	7	8	9	10
analýza As	0,2	0,4	0,0	0,9	0,3	0,1	0,0	0,2	0,2	0,1

(0040.xls)

Blok dat byl vygenerován generátorem náhodných čísel rovnoměrně rozložených. Posuďte rovnoměrnost rozložení sestrojením histogramu souboru dat a vypočtěte střední hodnotu a rozptyl tohoto souboru.

Považujte každý řádek definiční tabulky dat za výběr z tohoto souboru, určete u každého výběru střední hodnotu.

Určete i střední hodnotu a rozptyl souboru těchto výběrových průměrů. Pro tento soubor zkonstruujte také histogram.

(zadání 0044.xls)

Pro statistický soubor daný v tabulce určete základní statistické charakteristiky a ověřte, zda mohl být vybrán ze základního souboru normálně rozloženého.

53,0	79,7	71,4	84,0	74,7	76,4	68,7	58,9	87,6	96,4	60,3
82,8	70,3	49,3	99,1	75,7	59,2	73,3	57,9	87,1	46,7	100,7
67,7	42,8	49,0	63,0	90,0	46,6	65,9	43,8	86,4	80,3	57,3
45,5	52,7	69,9	68,0	65,9	62,1	87,1	70,8	85,3	68,1	63,4
73,5	62,6	77,4	76,3	45,1	61,9	83,5	45,6	88,8	47,4	69,6
86,8	81,1	57,4	67,5	86,4	71,1	87,6	46,1	71,3	74,6	90,3
104,9	67,2	79,3	67,3	77,5	43,8	82,3	44,2	99,0	69,4	58,1
75,6	58,8	66,9	96,6	65,9	68,1	87,7	82,3	86,1	85,8	58,6

87,2	51,1	76,6	39,6	85,5	41,6	42,6	70,5	41,9	101,8	72,8
79,4	46,1	90,4	78,2	76,8	63,1	54,7	83,2	53,0	58,0	60,7
48,8	74,1	61,4	43,6	82,0	70,7	60,4	61,7	70,4	56,9	61,3
51,9	86,4	73,8	83,6	62,2	76,7	65,5	46,6	42,8	25,6	79,4
43,8	96,2	41,2	82,4	83,8	51,2	48,1	40,3	76,1	69,0	58,9
64,7	62,1	80,4	68,7	71,2	47,2	64,5	84,2	67,3	46,7	63,0
66,2	74,8	74,6	72,4	62,4	63,8	60,4	46,7	48,0	42,1	68,9
75,8	69,7	79,5	56,5	44,6	95,7	84,7	43,9	45,1	99,6	41,1
55,4	35,5	57,1	79,7	66,4	79,6	80,6	59,8	81,0	74,3	83,6
82,5	47,2	63,7	69,2	66,7	88,9	77,5	68,0	65,5	76,2	62,7
95,1	65,2	72,2	90,7	62,5	48,3	72,6	66,5	70,4	59,5	80,0
61,5	82,7	94,1	42,7	62,8	65,6	65,6	101,4	63,7	58,7	44,7
84,6	59,7	53,9	78,3	89,6	86,5	44,3	74,0	46,4	73,4	97,8
59,0	55,6	41,1	101,2	90,8	60,8	117,2	68,2	67,2	82,1	84,6
40,3	68,0	71,1	68,7	76,6	74,0	70,4	61,1	51,0	45,3	79,4
81,9	71,9	53,8	69,7	90,5	49,5	82,2	62,2	54,5	64,1	47,5
67,0	37,3	76,5	43,2	60,2	50,0	79,7	94,6	85,3	44,8	91,8

(0045.xls)

Na stavbu byly dovezeny cihly ze tří cihelen a složeny na společné skládce. Jejich množství jsou v poměru 1:2:2. Cihly vyrobené jednotlivými cihelnami vyhoví předepsaným normám jakosti s pravděpodobností rovnou postupně 0,80, 0,65, 0,72. Ze skládky cihel náhodně vybereme jeden kus, abychom laboratorně zjistili, zda splňuje předepsané požadavky. Jaká je pravděpodobnost toho, že cihla bude mít předepsanou kvalitu?

(0046.xls)

K zvýšení spolehlivosti zařízení je blok **a** zdvojen (paralelní zapojení podle obrázku).

a) Když spolehlivost bloku **a** je $\boldsymbol{p}$, určete pravděpodobnost $\boldsymbol{P}$ celého zařízení a porovnejte se zařízením s jedním blokem. Proveďte pro různé hodnoty $\boldsymbol{p}$.

b) Řešte zvýšení spolehlivosti zařízení paralelním zapojením n bloků **a**.

c) Kolik je třeba zapojit bloků **a**, aby spolehlivost celého zařízení byla $\boldsymbol{P_1}$?

(0048.xls)

V městě byl po dobu 60 dnů evidován počet dopravních nehod v průběhu každého dne a podle počtu nehod v jednom dni vytvořena následující tabulka. Pro počet nehod v jednom dni jako náhodnou proměnnou sestrojit zákon rozložení, střední hodnotu a disperzi a ostatní momentové charakteristiky.

počet nehod / den	0	1	2	3	4	5	6
počet dnů s uvedeným počtem nehod	4	28	10	7	6	4	1

(0049.xls) (experimentální řešení viz 0073.xls)

Výsledkem náhodného pokusu je náhodná veličina, nabývající hodnot $\mathbf{1}/\boldsymbol{n}$ s pravděpodobnostmi nepřímo úměrnými $\mathbf{3}^{\boldsymbol{n}}$. Určete střední hodnotu a rozptyl této veličiny.

(0050.xls - řešení na listě 2)

Určete charakteristiky dvojrozměrných souborů včetně vhodné regresní funkce.

x	7	1	11	11	7	11	3	1	2	21	1	11	10
y	78,5	74,3	104,3	87,6	95,9	109,2	102,7	72,7	93,1	115,9	83,8	113,3	109,4

(0050.xls - řešení na listě 3)

x	5	9,6	16,0	19,6	24,4	29,8	34,4
y	2,60	2,01	1,34	1,08	0,94	1,06	1,25

(zadání 0050.xls)

x	0,0	0,5	1,0	1,5	2,0	2,5	3,0
y	0,0	1,7	3,1	3,8	3,9	3,8	3,0

(zadání 0050.xls)

x	55	65	75	85	95	105	115	125	135	145
y	1,74	2,02	2,12	2,05	2,17	2,47	2,4	2,48	2,5	2,39

x - délka stěny v rubání

y - produktivita

(zadání 0050.xls)

x	0,030	0,030	0,032	0,040	0,046	0,048	0,050
y	29,0	29,5	29,0	31,0	32,0	31,5	32,3

x - obsah síry v oceli(% S)

y - pevnost oceli v tahu (kg/mm^2)

(zadání 0050.xls)

x	65	66	67	68	69	70	71	72	73	74
y	69,2	70,1	71,0	71,8	72,7	73,6	74,5	75,4	76,2	77,1

75	76	77	78	79	80	81	82	83	84	85
78,0	78,9	79,8	80,6	81,5	82,4	83,3	84,2	85,0	85,9	86,8

x - výnos laboratorně stanovené neprchavé hořlaviny

y - provozní výnos koksu

(zadání 0050.xls)

obsah uhlíku v uhlí	90,5	89,0	88,6	91,3	90,0	87,5	86,8	86,0	84,6	84,6	88,8	87,0	86,7	83,9	87,6	84,7
součinitel melitelnosti	1,201	1,032	1,032	1,037	0,663	0,537	0,512	0,451	0,360	0,340	0,840	0,603	0,410	0,439	0,375	0,426

(zadání 0050.xls)

x	34,9	34,4	28,5	23,7	19,6	24,3	29,2	27,1	32,5	33,3	34,2	28,4	29,3	17,3	22,2	24,9	27,6	29,4	19,8	24,5	29,8	26,2
y	69,3	69,7	74,9	79,1	82,8	78,6	74,3	76,2	71,4	70,7	69,9	75,0	74,2	84,8	80,5	78,0	75,7	74,1	82,6	78,4	73,8	76,9

x - obsah prchavé hořlaviny v hořlavině uhlí (% hmotnosti)

y - provozní výnos koksu (% hmotnosti)

(zadání 0050.xls)

x	18,45	23,86	24,77	13,36	14,84	29,37	28,79	32,99	32,11	34,57	25,74	28,17	32,21	1,59	33,07	34,11
y	1,84	1,87	1,96	2,06	3,03	3,04	3,11	5,14	6,22	6,44	3,46	4,61	4,56	5,77	5,73	8,85

x - obsah prchavé hořlaviny v uhlí

y- součinitel melitelnosti

(zadání 0050.xls)

x	0,803	0,874	0,782	1,050	1,050	1,120	0,996	0,867	0,844	0,965
y_1	67,7	72,4	63,2	82,8	81,6	83,3	64,2	66,5	44,5	70,7
y_2	12,8	8,0	9,1	5,8	5,5	5,3	8,4	11,4	10,6	11,3

x - koksotvorný faktor G

y_1 - pevnostní ukazatel koksu M 40

y_2 - pevnostní ukazatel koksu M 10

(zadání 0050.xls)

C^{daf} %	90,54	89,03	88,61	91,33	90,03	87,52	86,80	86,02
v^{daf} %	18,45	23,86	24,77	13,36	14,84	29,37	28,79	32,99
A	1,84	1,87	1,96	2,06	3,03	3,04	3,11	5,14

84,55	84,55	88,82	86,98	86,68	83,89	87,61	84,71
32,11	31,57	25,74	28,17	32,21	31,59	33,07	34,11
6,22	6,44	3,46	4,61	4,56	5,77	5,73	8,85

C - obsah uhlíku v uhlí

v - množství prchavé hořlaviny v uhlí

A - práce potřebná k drcení uhlí

(zadání 0050.xls)

x	1,224	1,233	1,251	1,261	1,218	1,233	1,253	1,261	1,221	1,236	1,250	1,263
y	0,45	0,89	1,44	1,98	0,42	0,95	1,46	2,00	0,43	0,93	1,45	1,99

x - A - vynaložená práce na drcení uhlí

y - obsah podsítného D 88 (pod 88 μm)

(zadání 0050.xls)

x	154	133	58	145	94	113	86	121	119	112	85	41	96	45	47
y	178	164	75	161	107	141	97	127	138	125	97	72	113	89	61
z	59	63	36	62	48	64	44	57	62	51	45	45	51	41	36
x	99	51	101	169	87	88	83	106	92	85	112	98	103	99	68
y	109	95	114	209	101	139	98	111	104	103	118	102	108	119	85
z	49	46	63	73	55	65	46	58	45	46	55	48	50	60	38
x	104	107	98	97	105	71	39	122	33	78	114	125	73	77	137
y	128	118	140	115	101	93	69	147	52	117	138	149	76	85	142
z	41	65	40	66	55	43	30	55	25	56	62	63	32	43	61
x	44	92	141	155	136	82	136	72	66	42	113	42	133	153	85
y	69	116	157	193	155	81	163	79	81	61	123	85	147	179	91
z	32	48	54	60	65	41	85	43	40	29	49	36	52	72	48

vlastnosti oceli:

x - mez tahu (kp/mm^2)

y - pevnost v lomu (kp/mm^2)

z - mez pružnosti (kp/mm^2)

(0051.xls)

Údaje o prodeji chladniček určitého typu za roky 1971 - 1985 vyrovnejte logistickou křivkou.

rok	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
y	25	50	90	180	280	800	1 460	2 700	4 800	7 600	11 100	14 200	16 800	17 600	18 400

(zadání 0052.xls)

Určete základní charakteristiky následujících časových řad

rok	1980	1981	1982	1983	1984	1985
výroba plynu (m^3)	1286	1363	1393	1495	1571	1610

(zadání 0052.xls)

měsíc (1985)	1	2	3	4	5	6	7	8	9	10	11	12
plánovaná těžba (t)	41000	40000	43000	44000	44000	42000	40000	40000	42000	44000	45000	45000
skutečná těžba (t)	42605	38690	45694	43122	39526	39636	37765	35813	42265	49711	49089	47030

(zadání 0052.xls)

rok	1977	1978	1979	1980	1981	1982	1983	1984	1985
y	37,5	39,3	41,4	42,9	45,1	47,2	49,6	51,2	53,4

y - velikost výroby membránových filtrů (v tisících kusů)

Předpokládejte, že není dosud známá hodnota výroby v roce 1985. Zkuste na základě předešlých výsledků odhadnout tuto hodnotu extrapolací vhodné regresní funkce.

(zadání 0052.xls)

rok	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
výroba el. energie (tis. kWh)	5,6	6,7	7,5	8,3	9,3	10,3	11,6	12,4	13,6	15,0	16,6

(zadání 0052.xls)

rok	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
spotřeba mražených jídel (ve 100 kg)	133	155	195	361	310	373	618	1 108	1 263	1 600	2 172	2 563	3 202	3 892	3 964	4 600	5 100	5 461

(0053.xls) (zadání 0052.xls)

rok	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
vyrobeno traktorů	2986	5010	7355	7532	8473	8910	10021	10479	10523	10754	10950	11121

(modifikovaná trendová exponenciální křivka)

(zadání 0052.xls)

Průměrný věk nevěst a ženichů (zdroj: ČSÚ)

rok	1991	1993	1995	1996	1997	1998	1999	2000	2001
muži	24,7	25,4	26,7	27,1	27,6	28,1	28,5	28,8	29,0
ženy	22,2	23,2	24,6	24,9	25,4	25,7	26,2	26,4	26,9

(zadání 0052.xls)

rok	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
počet svateb v ČR	90953	71937	74060	66033	58440	54956	53896	57804	55027	53523	55321	52374	53732

(zadání 0054.xls)

Byly měřeny dvě vlastnosti litiny sig a sig2 a provedena chemická analýza složení vzorků. Posuďte, která složka nejvíce ovlivňuje sledované vlastnosti a změřte jejich přínos.

C	Zr	Ti	sig	sig2
0,0267	0,2491	0,1639	62,4691	79,5995
0,0597	0,1488	0,3083	73,8822	73,5017
0,0628	0,1716	0,2375	78,8197	79,2880
0,0018	0,0546	0,2608	71,3198	57,5080
0,0368	0,1576	0,3656	82,0695	71,5656
0,0016	0,2485	0,3572	86,7472	91,7285
0,0739	0,2696	0,2674	102,3706	90,6495
0,0042	0,0019	0,2555	99,2234	96,7699
0,0599	0,2473	0,2900	76,3294	77,1619
0,0479	0,1543	0,2945	85,4812	66,5626

0,0768	0,1453	0,2011	69,6071	90,7690
0,0398	0,1691	0,3133	95,2214	66,3793
0,0547	0,0805	0,1749	77,3614	71,0235
0,0368	0,0706	0,3869	81,4018	69,2754
0,0422	0,1075	0,2395	78,0598	70,4878
0,0679	0,2158	0,2767	100,3271	85,4372
0,0152	0,0992	0,2968	85,2486	96,3644
0,0457	0,0398	0,3037	84,1396	74,3663
0,0582	0,1008	0,3421	92,9368	68,9465
0,0535	0,1124	0,2936	70,9373	84,7529
0,0815	0,1820	0,2376	80,1945	62,6996
0,0415	0,2731	0,1672	89,4634	71,4948
0,0412	0,1894	0,1887	79,2855	79,3510
0,0246	0,1708	0,3360	67,3449	73,1299
0,0152	0,1265	0,2675	67,4148	63,5108

(0055.xls)

Posuďte vliv jednotlivých vybraných ukazatelů parních elektráren v roce 1984 na měrné náklady elektráren. Úlohu řešte vicenásobnou lineární reresní analýzou.

elektrárna	měrné náklady (Kč/MWh)	poruchy (%)	využití pohotového výkonu (tisíce hodin)	cena paliva (Kč/GJ)	měrná spotřeba (GJ/MWh)
	y	x_1	x_2	x_3	x_4
Mělník 2	249	0,95	6,86	14,01	12,92
Počerady 1	203	2,27	7,56	12,06	11,74
Chvaletice	256	2,34	6,79	15,03	11,74
Dětmarovice	306	4,34	7,25	17,38	11,7
Tušimice 1	227	2,22	6,58	10,28	12,49
Tušimice 2	213	2,62	7,35	10,12	12,13
Prunéřov 1	349	5,18	6,66	11,26	13,49
Prunéřov 2	210	4,24	7,47	11,53	11,15

(0056.xls)

Určete lineární regresní funkci pro data (x, y) v tabulce. Pokuste se tento lineární model vylepšit pro účely extrapolace pro větší hodnoty x tím, že zavedete váhy jednotlivých bodů (body s větší x-ovou souřadnicí mají větší váhu).

x	1	2	3	4	5
y	1	3	4	4	5

(0057.xls)

Otestujte, zda u dvojrozměrné veličiny dané v tabulce může jít o lineární závislost.

x	0,0	0,5	1,0	1,5	2,0	2,5	3,0
y	0,0	1,7	3,1	3,8	3,9	3,8	3,0

(0075.xls)

Sledujte průběh funkce binomického rozložení náhodné veličiny.

Srovnejte s průběhem vhodné funkce Poissonova a normálního rozložení.

(zadání 0076.xls)

Při stavbě betonové konstrukce bylo odebráno 100 vzorků betonové směsi. Po 28 dnech (stanoveno normou) vykázaly zkušební kostky tuto krychelnou pevnost (kp/cm^2):

270	247	214	249	282	309	272	250	219	226
270	323	254	277	256	260	238	231	251	310
272	221	189	295	182	267	270	253	222	225
206	303	253	256	281	232	230	186	200	252
222	279	256	229	316	275	216	245	197	266
265	241	296	176	273	245	310	224	252	276
198	232	238	256	286	291	257	232	236	256
277	287	225	196	291	268	266	243	263	247
263	237	260	281	282	259	230	210	240	242
235	305	297	269	244	262	238	260	246	262

Vypočtěte výběrové charakteristiky a rozhodněte, zda vzorek pochází ze souboru normálně rozloženého.

Ve středoškolských učebnicích z různých předmětů (Čj, D, Bi, F) byly sledovány počty vět ve větných celcích. Výsledky v tabulce:

počet vět	1	2	3	4	5	6	7	8	9	10
Čj	753	421	163	70	39	3	2	0	0	1
D	1459	978	355	71	12	5	1	1	0	0
Bi	1317	718	206	36	12	1	2	0	0	0
F	1604	1289	583	124	32	7	4	2	0	0

Zpracujte tyto údaje statisticky a zformulujte otázky, na které by mohla odpovědět statistická indukce.

(0077.xls)

Při seskoku parašutisty byla měřena závislost mezi rychlostí v a tlakem p na povrch padáku. Výsledky vyrovnejte parabolou $p = a + b.v^2$.

v m/s	2,40	3,50	5,00	6,89	10,00
p 0,1 mPa	0,0141	0,0281	0,0562	0,1125	0,2250

Závislost mezi cenou žita, jako měřítka ceny nejnutnější životní potřeby širokých vrstev lidových a poměrnou četností přestupků krádeže, jako měřítka kriminality těchto vrstev (citace: Prof. Dr. Cyril Horáček ml.: Úvod do studia statistiky, Nákladem Spolku československých právníků "Všehrd" 1932)

rok	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898
cena žita v markách za 100 kg	180	164	154	152	143	143	157	170	182	215	185	141	122	134	138	154	171
počet přestupků krádeže na 100 000 obyvatel	250	239	230	210	210	196	190	210	205	215	234	200	196	191	181	188	194

(0078.xls - studentská práce s připomínkami učitele)

Pro výrobu drátu se používají tři jakosti vstupní suroviny. V laboratoři byly naměřeny pevnosti (v MPa) již vyrobeného drátu. Posuďte významnost rozdílů a výběrových průměrů mezi jednotlivými jakostmi. (Data viz citovaný sešit excelu.)

(0079.xls - studentská práce)

Posuďte vliv jednotlivých prvků na množství přetrhů během tažení drátu pro různé jakosti válcovaného drátu (A-G).

	Přetrhy (1/100 t)	%C	%Mn	%Si	%P
A	80	0,05	0,15	0,45	0,004
B	75	0,08	0,2	0,33	0,002
C	78	0,07	0,11	0,32	0,002
D	65	0,04	0,12	0,36	0,003
E	45	0,03	0,13	0,35	0,004
F	72	0,08	0,15	0,35	0,005
G	75	0,07	0,19	0,45	0,007

(0081.xls - studentská práce)

Počet obyvatel k 1.7.1994 podle věku

věková skupina	0	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39
muži	57 969	256 287	333 344	366 536	458 571	407 149	350 709	335 273	369 257
ženy	55 074	243 050	317 880	348 862	439 712	388 419	335 923	322 958	362 492

40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
408 768	398 013	306 376	229 692	232 719	203 940	158 759	63 820	58 945	25 281
406 847	403 006	319 460	254 288	276 623	276 810	249 295	115 111	126 213	72 731

Počet obyvatel k 1.7.1994 podle regionů

region	PRAHA	StČ	JhČ	ZpČ	SvČ	VchČ	JhM	SvM
muži	573 079	540 437	343 788	421 603	575 362	602 933	1 000 207	963 999
ženy	643 489	568 256	356 900	440 355	602 790	634 474	1 058 852	1 009 638

(Zkuste vytěžit z těchto dat více, než nabízí řešení v sešitě 0081.xls.)

V karetní hře SRDCE, kterou nabízí OS Windows, hraje uživatel počítače (hráč A) proti třem soupeřům, kteří reprezentuji počítač (hráči PC1, PC2, PC3).

Po 150 partiích (partie končí,, když aspoň jeden hráč získá aspoň 100 trestných bodů, vítězí pak ten, kdo získá nejméně trestných bodů) bylo zjištěno, že

a) počet vyhraných partií je pro jednotlivé hráče dán vektorem v = (A, PC1, PC2, PC3) = (51, 31, 32, 36),

b) součet získaných trestných bodů je dán vektorem b = (A, PC1, PC2, PC3) = (10285, 11 531, 11 708, 11 312).

Vyjádřete se k úrovni hry hráče A vzhledem ke hře jeho soupeřů PC1, PC2, PC3.

(zadání 0082.xls)

Jsou známy bodové výsledky zkouškového testu u čtyř stejně početných skupin studentů:

	interval hodnot získaných bodů														
skupina studentů	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169
1	1	4	6	8	10	16	18	16	10	8	6	4	1	0	0
2	0	2	5	10	16	17	18	12	10	7	5	3	1	1	1
3	0	0	12	12	12	12	12	12	12	12	12	0	0	0	0
4	0	0	0	34	12	6	4	6	12	34	0	0	0	0	0

Určete základní statistické ukazatele pro každou skupinu studentů.

(viz citovaná literatura Hanousek, Chamrada, str. 38n.)

Zkouškami bylo zjištěno, že střední doba životnosti určitého typu elektronek je 1250 hodin. Doba životnosti se řídí exponenciálním rozdělením.

a) Jaká je pravděpodobnost, že náhodně vybraná elektronka bude mít životnost kratší než 500 hodin?

b) Jaká je pravděpodobnost, že náhodně vybraná elektronka bude mít životnost delší než 2000 hodin?

c) Jaká je pravděpodobnost, že náhodně vybraná elektronka nebude mít větší odchylku od střední doby životnosti než 100 hodin?