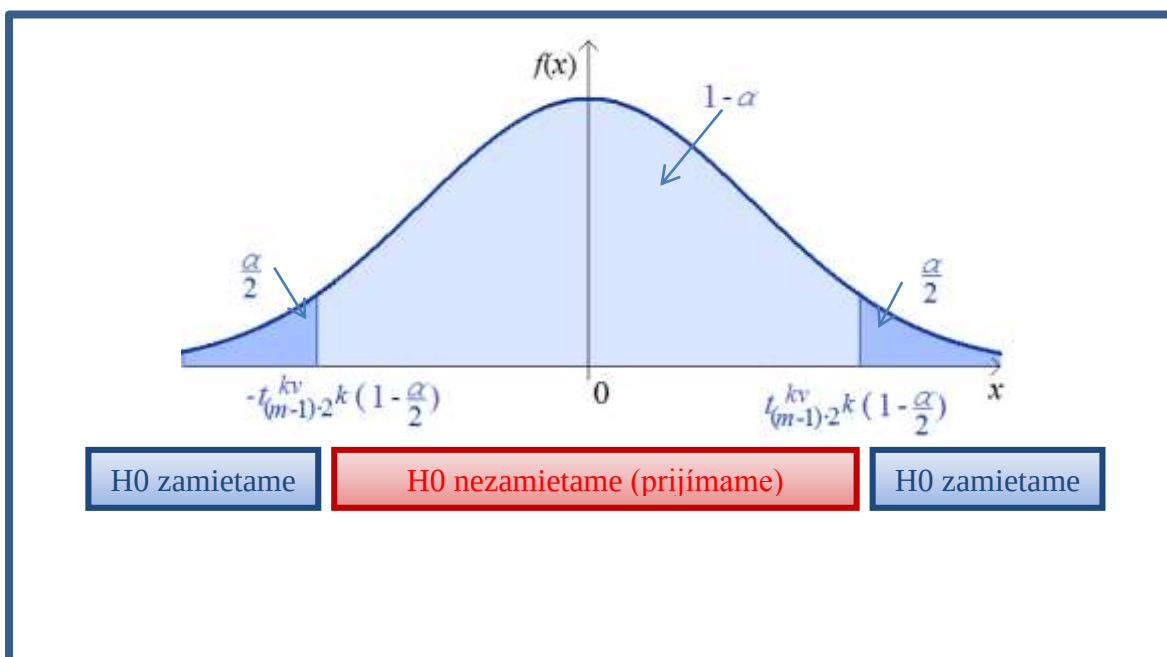


MATEMATICKÉ METÓDY PLÁNOVANIA A VYHODNOCOVANIA EXPERIMENTU

TÓTHOVÁ Mária
KOTIANOVÁ Janette



2015

© Materiálovotechnologická fakulta STU so sídlom v Trnave

RNDr. Mária Tóthová, PhD., PaedDr. Janette Kotianová, PhD.,

Recenzenti: doc. Ing. Dušan Holý, CSc.
doc. RNDr. Oleg Palumbíny, CSc.

Jazyková korektúra: Mgr. Mária Mičeková

Schválilo Vedenie Materiálovotechnologickej fakulty STU ako vysokoškolské skriptá dňa 25. novembra 2014 pre študijné programy Materiálové inžinierstvo, Počítačová podpora návrhu a výroby, Obrábanie a montáž, Zváranie, Výrobné zariadenia a systémy, Priemyselné manažérstvo Materiálovotechnologickej fakulty STU v Trnave

ISBN 978-80-8096-227-2

EAN 9788080962272

PREDHOVOR

Táto publikácia je určená študentom prvého ročníka inžinierskeho štúdia na MTF STU v Trnave ako študijný text k predmetu MATEMATICKÉ METÓDY PLÁNOVANIA A VYHODNOCOVANIA EXPERIMENTU. Svojím obsahom nadväzuje na skriptá s rovnomenným názvom z roku 2009, pričom reaguje na reorganizáciu cvičení z predmetu. Pôvodný obsah bol doplnený o pojem diskkrétnej a spojitej náhodnej premennej, základné typy rozdelení, základy teórie pravdepodobnosti a kombinatoriku, ktoré umožňujú študentom lepšie chápať súvislosti a interpretovať výsledky výpočtov.

V snahe priblíženia predmetu MMPVE s praxou bolo prirodzené do výučby zaradiť softvér. Ťažisko cvičení tak nie je v ručnom numerickom spracovávaní údajov, ale v správnom výbere ponúkaných funkcií, v korektnom zadaní vstupov a v interpretácii výstupov softvéru. Napriek tomu, že existuje množstvo špeciálnych programov, ktoré poskytujú komplexnú štatistickú analýzu dát, na výučbu základov štatistiky a regresnej analýzy, ktoré sú súčasťou predmetu, bol vybraný ako postačujúci tabuľkový procesor Excel, ktorý je dostupný každému užívateľovi programového balíka MS Office a študenti väčšinou základy práce v ňom ovládajú.

Obsah predkladaných učebných textov je rozdelený do 5 kapitol a časti obsahujúcej prílohy. Každá kapitola obsahuje okrem stručného vymedzenia základných pojmov aj riešené príklady s podrobným popisom ich riešenia v Exceli a s interpretáciou výsledkov. Kvôli názornosti sú pri definovaných pojmoch a riešených príkladoch vyžadujúcich u študenta predstavu riešeného problému uvedené grafické zobrazenia pojmov a riešených situácií. V závere sú vždy uvedené neriešené úlohy s výsledkami. Postupy a návody riešení sú vo verzii Excelu 2010, ktorú študenti na cvičeniach aktuálne využívajú. Súčasťou publikácie sú prílohy, ktoré obsahujú v Exceli riešené úlohy vybraných častí základného kurzu matematickej analýzy, tiež popis spustenia doplnku Analytické nástroje, ktorý nie je štandardnou súčasťou programu Excel a metódu najmenších štvorcov, ktoré sa využívajú pri analýze dát a pri riešení úloh regresnej analýzy.

Publikácia bola písaná s myšlienkou uľahčiť študentom pochopenie rozsiahleho a náročného učiva a pomôcť im zorientovať sa v oblasti teórie pravdepodobnosti a štatistiky a taktiež v teórii plánovania a vyhodnocovania experimentov.

Autori touto cestou ďakujú recenzentom doc. RNDr. Olegovi Palumbínymu, CSc. a doc. RNDr. Dušanovi Holému, CSc. za pripomienky, ktoré pomohli k skvalitneniu odborného obsahu predkladaných učebných textov.

1. KOMBINATORIKA

Pri riešení úloh z teórie pravdepodobnosti často používame základné vzorce z kombinatoriky, t.j. náuky o permutáciach, variáciach a kombináciach.

Nech množina $M = \{a_1, a_2, \dots, a_n\}$ je množina n ($n \in \mathbb{N}$) navzájom rôznych prvkov. Pod „výberom“ rozsahu k ($k \in \mathbb{N}$) budeme rozumieť množinu, vytvorenú z k prvkov množiny M (množinu „vybraných“ k prvkov z n prvkovej množiny). Ak výber realizujeme s opakovaním (prvok množiny M sa môže dostať do výberu viackrát), hovoríme o výbere s opakovaním, v opačnom prípade hovoríme o výbere bez opakovania. Ak výbery, ktoré sa líšia iba usporiadaním prvkov budeme rozlišovať, hovoríme, že záleží na poradí, v opačnom prípade hovoríme, že nezáleží na poradí.

Definícia 1.1 Výbery s rozsahom k bez opakovania ($k \leq n$), pri ktorých záleží na poradí prvkov, nazývame variácie bez opakovania. Ak $k = n$ výbery sa nazývajú permutácie bez opakovania.

Definícia 1.2 Výbery s rozsahom k s opakovaním (k môže byť aj väčšie ako n), pri ktorých záleží na poradí prvkov, nazývame variácie s opakovaním.

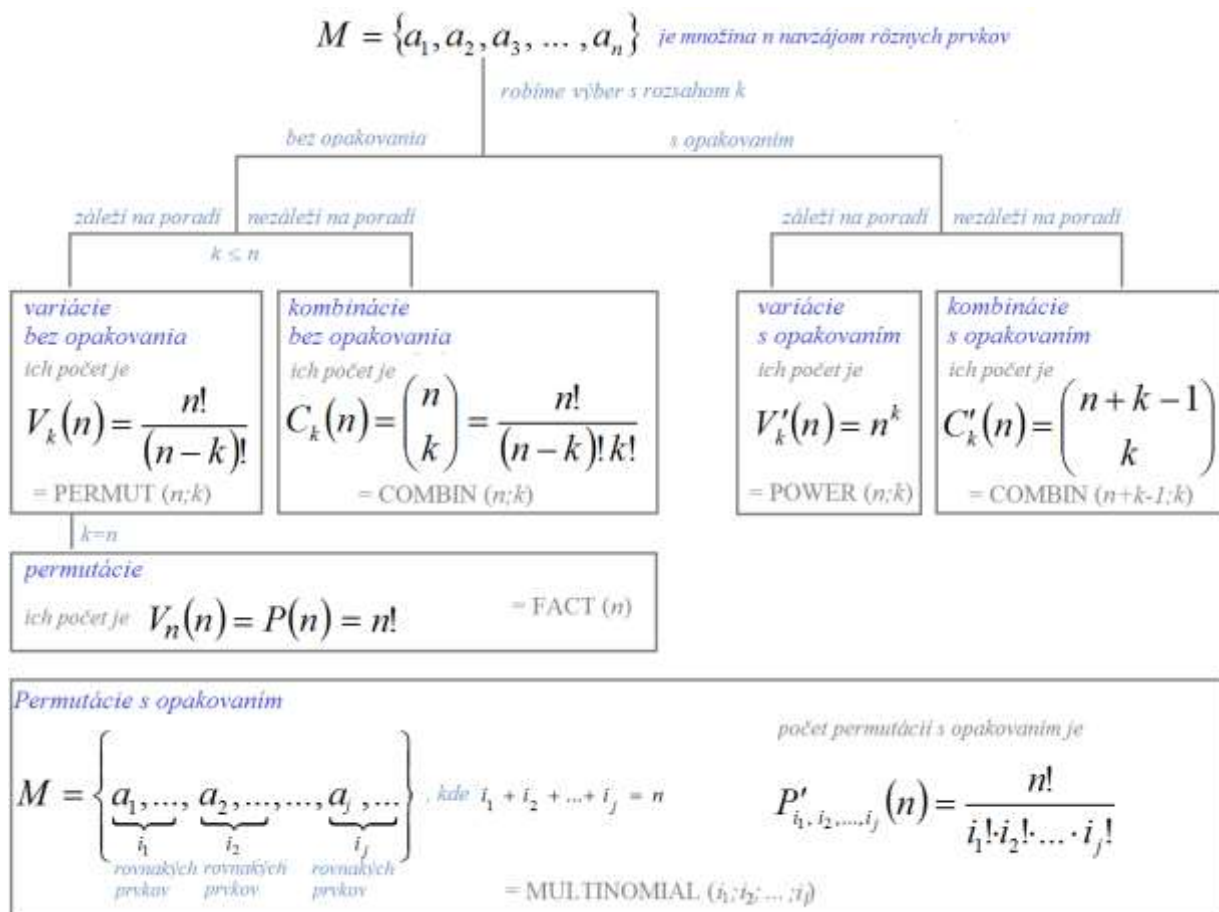
Definícia 1.3 Výbery s rozsahom k bez opakovania ($k \leq n$), pri ktorých nezáleží na poradí prvkov, nazývame kombinácie bez opakovania.

Definícia 1.4 Výbery s rozsahom k s opakovaním (k môže byť aj väčšie ako n), pri ktorých nezáleží na poradí prvkov, nazývame kombinácie s opakovaním.

Definícia 1.5 Nech množina M obsahuje i_1 rovnakých prvkov a_1 , i_2 iných rovnakých prvkov a_2 , ..., i_j opäť iných rovnakých prvkov a_j , kde $i_1 + i_2 + \dots + i_j = n$. Výbery s rozsahom $k = n$ nazývame permutácie s opakovaním.

Poznámka 1.1 Kombinácie sú neusporiadané výbery, variácie sú usporiadané výbery a permutácie sú špeciálnym prípadom variácií.

Počet kombinácií, variácií, permutácií a ich výpočet aj pomocou funkcií v Exceli sú uvedené v nasledujúcej schéme:



Poznámka 1.2 Verzia Excelu 2013 má na výpočet variácie s opakovaním príkaz =PERMUTATIONA($n; k$) = n^k .

Príklad 1.1 Daná je množina $M = \{a, b, c, d\}$. Z prvkov tejto množiny vytvorte kombinácie 2. triedy bez opakovania a kombinácie s opakovaním. Vypočítajte ich počet.

Riešenie. Hľadáme výbery s rozsahom $k = 2$, nezáleží na poradí prvkov, najprv bez opakovania a potom s opakovaním prvkov.

Kombinácie bez opakovania:

$$C_2(4): \left\{ \begin{array}{l} ab, ac, ad, \\ bc, bd, \\ cd \end{array} \right\}$$

Ich počet je

$$C_2(4) = \binom{4}{2} = \frac{4!}{2!2!} = \text{COMBIN}(4; 2) = 6.$$

Kombinácie s opakovaním:

$$C'_2(4): \left\{ \begin{array}{l} aa, ab, ac, ad, \\ bb, bc, bd, \\ cc, cd, \\ dd \end{array} \right\}$$

Ich počet je

$$C'_2(4) = \binom{5}{2} = \frac{5!}{2!3!} = \text{COMBIN}(5; 2) = 10.$$

Príklad 1.2 Daná je množina $M = \{a, b, c, d\}$. Z prvkov tejto množiny vytvorte variácie 2. triedy bez opakovania a s opakovaním. Vypočítajte ich počet.

Riešenie. Vytvárame výbery s rozsahom $k = 2$, záleží na poradí prvkov, bez opakovania a následne s opakovaním prvkov.

Variácie bez opakovania:

$$V_2(4): \left\{ \begin{array}{l} ab, ac, ad, \\ ba, bc, bd, \\ ca, cb, cd, \\ da, db, dc \end{array} \right\}$$

Ich počet je

$$V_2(4) = \frac{4!}{2!} = \text{PERMUT}(4;2) = 12.$$

Variácie s opakovaním:

$$V'_2(4): \left\{ \begin{array}{l} aa, ab, ac, ad, \\ ba, bb, bc, bd, \\ ca, cb, cc, cd, \\ da, db, dc, dd \end{array} \right\}$$

Ich počet je

$$V'_2(4) = 4^2 = \text{POWER}(4;2) = 16.$$

Príklad 1.3 Daná je množina $M = \{a, b, c\}$. Utvorte všetky permutácie tretej triedy z prvkov tejto množiny a vypočítajte ich počet.

Riešenie. Permutácie tretej triedy sú rôzne usporiadania prvkov tejto množiny, t.j. výbery s rozsahom $k = 3$, bez opakovania a záleží na poradí prvkov.

$$P(3): \{abc, acb, bac, bca, cab, cba\}$$

Ich počet je $P(3) = 3! = \text{FACT}(3) = 6$.

Príklad 1.4 Daná je množina $M = \{a, a, c, d\}$. Vytvorte z tejto množiny všetky permutácie s opakovaním.

Riešenie. Množina M neobsahuje navzájom rôzne prvky, prvé dva sú rovnaké. Permutácie s opakovaním znamenajú rôzne možné usporiadania prvkov tejto množiny, ktoré budú $\{aacd, aadc, caad, daac, cdaa, dcaa, \\ acad, adac, acda, adca, cada, daca\}$.

Ich počet je $P'_{2,1,1}(4) = \frac{4!}{2!1!1!} = \text{MULTINOMIAL}(2; 1; 1) = 12$.

Príklad 1.5 Koľkými spôsobmi možno vytvoriť výrobnú linku pozostávajúcu z 5 rôznych montážnych panelov?

Riešenie. Množinu M (z čoho „vyberáme“) tvorí 5 navzájom rôznych montážnych panelov.

$M = \{5 \text{ panelov} \}; n = 5$.

Výber sa uskutočňuje s rozsahom $k = 5$, bez opakovania a záleží na poradí prvkov. Sú to permutácie bez opakovania, ich počet je

$$P(5) = 5! = \frac{5!}{(5-5)!} = \text{FACT}(5) = 120.$$

Príklad 1.6 Koľko rôznych dvojprvkových prevodov (dve ozubené kolesá) možno vytvoriť zo sady 8 ozubených koliesok o rôznom počte zubov?

Riešenie. Množinu M tvorí 8 navzájom rôznych ozubených koliesok.

$M = \{8 \text{ koliesok}\}$; $n = 8$. Pre vytvorenie prevodov vyberáme 2 kolieska. Ide o výber s rozsahom $k = 2$, výber bez opakovania a záleží na poradí kolies. Dostaneme variácie bez opakovania, ich počet je

$$V_2(8) = \frac{8!}{(8-2)!} = \text{PERMUT}(8;2) = 56.$$

Príklad 1.7 Koľko rôznych výsledkov (dvojíc bodov) môžeme hodiť pri hode 2 rôznymi hracími kockami?

Riešenie. Na kocke môže padnúť počet bodov 1 alebo 2 alebo 3 alebo 4 alebo 5 alebo 6. $M = \{1,2,3,4,5,6\}$; $n = 6$

Rozsah výberu je $k = 2$, výber sa realizuje s opakovaním a záleží na poradí. Ide o variácie s opakovaním, ktorých počet je

$$V'_2(6) = 6^2 = \text{POWER}(6;2) = 36.$$

Príklad 1.8 Koľkými spôsobmi je možné vybrať z dvanásťčlenného tímu trojčlennú pracovnú skupinu.

Riešenie. Základnú množinu M tvoria členovia tímu.

$M = \{12 \text{ členov}\}$; $n = 12$, vyberáme $k = 3$, bez opakovania a nezáleží na poradí členov.

Budú to kombinácie bez opakovania, ktorých počet vypočítame ako

$$C_3(12) = \binom{12}{3} = \text{COMBIN}(12;3) = 220.$$

Príklad 1.9 Koľko je možností kúpy 10 zväracích elektród na hliník, ak v obchode majú 5 rôznych druhov ALCORD12Si, ALCORD 5Si, ALCORD Al, ESAB OK 96.20, ESAB OK 96.40 (v dostatočnom množstve)?

Riešenie. Množina M sa skladá z 5 navzájom rôznych zväracích elektród.

$M = \{\text{ALCORD 12Si, ALCORD 5Si, ALCORD Al, ESAB OK 96.20, ESAB OK 96.40}\}$; $n = 5$

Robíme výber s rozsahom $k = 10$, s opakovaním a nezáleží na poradí vybraných prvkov. Máme kombinácie s opakovaním, ich počet je

$$C'_{10}(5) = \binom{5+10-1}{10} = \text{COMBIN}(5+10-1;10) = 1001.$$

Príklad 1.10 Šesťciferný kód uzatvárania trezoru je vytvorený z tých istých číslíc ako číslo 257225. Koľko je možností vytvorenia príslušného kódu?

Riešenie. Všetkých čísl je 6, z nich sú $i_1 = 3$ rovnaké ($a_1 = 2$), $i_2 = 2$ iné rovnaké ($a_2 = 5$) a $i_3 = 1$ opäť iných rovnakých ($a_3 = 7$).

$M = \{a_1, a_1, a_1, a_2, a_2, a_3, \dots\}$; $n = 6$, vyberáme $k = 6$, záleží na poradí, čiže prvky množiny M , ktoré nie sú navzájom rôzne, usporiadame. Dostaneme permutácie s opakovaním, ich počet určíme ako

$$P'_{3,2,1}(6) = \frac{6!}{3!2!1!} = \text{MULTINOMIAL}(3,2,1) = 60.$$

Príklad 1.11 Máme 20 výrobkov, z ktorých je 15 dobrých a 5 chybných. Koľkými spôsobmi môžeme vybrať 10 výrobkov tak, aby medzi nimi bolo 7 výrobkov dobrých a 3 chybné?

Riešenie. Množinu M tvorí 15+5 výrobkov. Vytvárame výbery s rozsahom 7 z dobrých výrobkov a výbery s rozsahom 3 z chybných výrobkov. Výbery sa realizujú bez opakovania a nezáleží na poradí. Dostaneme kombinácie bez opakovania. Celkový počet všetkých možností výberu 10 výrobkov tak, aby medzi nimi bolo 7 výrobkov dobrých a 3 chybné, bude ich súčin.

$$C_7(15) \cdot C_3(5) = \binom{15}{7} \cdot \binom{5}{3} = \text{COMBIN}(15; 7) * \text{COMBIN}(5; 3) = 30030.$$

Príklad 1.12 V cukrárni majú 5 druhov zmrzliny. Otec chce pre svoju rodinu kúpiť 15 porcií. Koľkými spôsobmi môže zmrzlinu kúpiť?

$$\text{Riešenie. } C'_{15}(5) = \binom{5+15-1}{15} = \binom{19}{15} = \text{COMBIN}(19;15) = 3\,876.$$

Príklad 1.13 Koľkými spôsobmi sa dá kúpiť 5 rôznych čokolád z 15 rôznych druhov?

$$\text{Riešenie. } C_5(15) = \binom{15}{5} = \text{COMBIN}(15;5) = 3\,003.$$

Príklad 1.14 Koľkými spôsobmi sa dá kúpiť 15 čokolád z 5 rôznych druhov?

$$\text{Riešenie. } C'_{15}(5) = \binom{5+15-1}{15} = \binom{19}{15} = \text{COMBIN}(19;15) = 3\,876.$$

Príklad 1.15 Skúšajúci má pripravených 15 príkladov z aritmetiky a 10 príkladov z geometrie. Na písomku chce dať 4 aritmetické a 1 geometrický príklad. Koľko má možností zostavenia rôznych zadaní?

Riešenie. Skúšajúci má možnosť vybrať z 15 aritmetických príkladov 4 príklady $C_4(15) = \binom{15}{4}$

spôsobmi a geometrický príklad môže vybrať z 10 geometrických príkladov $C_1(10) = \binom{10}{1}$ spôsobmi.

Počet zostavení príkladov je súčin

$$\binom{15}{4} \cdot \binom{10}{1} = \text{COMBIN}(15;4) * \text{COMBIN}(10;1) = 13650.$$

Príklad 1.16 V rovine je 6 rôznych bodov, pričom žiadne 3 neležia na jednej priamke. Koľko rôznych priamok dostaneme spájaním všetkých týchto bodov navzájom?

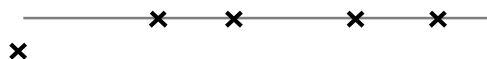
Riešenie. $C_2(6) = \binom{6}{2} = \text{COMBIN}(6;2) = 15.$

Pri spájaní bodov na poradi nezáleží (t.j. spojením bodu A a B dostaneme rovnakú priamku ako spojením bodov B a A).

Príklad 1.17 Koľkými priamkami možno spojiť 10 rôznych bodov v rovine, keď 4 ležia na jednej priamke a z ostatných už žiadne 3 neležia na tej istej priamke?

Riešenie.

Všetkých priamok je: $C_2(10) = \binom{10}{2} = 45.$



Totožné priamky: $C_2(4) = \binom{4}{2} = 6.$



Počet rôznych priamok: $C_2(10) - C_2(4) + 1 = 45 - 6 + 1 = 40.$



Príklad 1.18 V skupine 20 študentov, kde každý má iné meno, sú aj dievčatá Alena a Barbora. Koľkými spôsobmi môžeme zo študentov vybrať 8-člennú skupinu, aby medzi vybranými:

- bola Alena,
- nebola Alena,
- bola Alena aj Barbora,
- nebola Alena ani Barbora,

- e) bolo aspoň jedno z dievčat,
f) bolo najviac jedno z dievčat.

Riešenie. Ak označíme 20-člennú skupinu napr. A B C D E F G H I J K L M N O P R S T U, potom 8-člennú skupinu vyberáme nasledovnými spôsobmi:

a) Ak medzi 8 vybranými je Alena: A
 $C_7(19) = \binom{19}{7} = 50\,388;$ zvyšných 7 študentov vyberáme už iba z 19 študentov

b) Ak medzi 8 vybranými nemá byť Alena, musíme ju z 20-člennej skupiny na začiatku vylúčiť, t.j. 8 členov vyberáme z 19-tich:
 $C_8(19) = \binom{19}{8} = 75\,582;$

c) Ak chceme, aby medzi vybranými bola Alena aj Barbora: A B
 $C_6(18) = \binom{18}{6} = 18\,564;$ 6 študentov vyberáme z 18 študentov

d) Ak chceme, aby medzi vybranými nebola Alena ani Barbora:
 $C_8(18) = \binom{18}{8} = 43\,758;$ všetkých 8 študentov vyberáme z 18 študentov, t.j. A aj B sme zo základnej 20-člennej skupiny, z ktorej 8 študentov vyberáme, „vylúčili“

- e) Ak medzi vybranými má byť aspoň jedno z dievčat, môžu nastať možnosti:

- bude tam A a súčasne nebude B: A alebo
7 študentov vyberáme už iba z 18 študentov – A sme „minuli“ a B sme „vylúčili“

$C_7(18) = \binom{18}{7} = 31\,824;$

- bude tam B a súčasne nebude A: B alebo
7 študentov vyberáme už iba z 18 študentov – B sme „minuli“ a A sme „vylúčili“ – rovnaký princíp ako v predchádzajúcom prípade

$C_7(18) = \binom{18}{7} = 31\,824;$

- vo vybranej skupine bude aj A a súčasne aj B: A B
6 študentov vyberáme z 18 študentov

$C_6(18) = \binom{18}{6} = 18\,564;$

Počet všetkých možností je $C_7(18) + C_7(18) + C_6(18) = 82\,212$.

f) Ak medzi vybranými má byť najviac jedno z dievčat, buď:

- bude tam A a súčasne nebude B: $C_7(18) = \binom{18}{7} = 31\,824$;

- bude tam B a súčasne nebude A: $C_7(18) = \binom{18}{7} = 31\,824$;

- vo vybranej skupine nebude ani A a súčasne ani B: $\underbrace{\hspace{2cm}}$

$$C_8(18) = \binom{18}{8} = 43\,758;$$

8 študentov vyberáme z 18 študentov

Spolu počet všetkých možností bude $C_7(18) + C_7(18) + C_8(18) = 107\,406$.

Príklad 1.19 Tréner futbalového mužstva má 3 brankárov, 5 obrancov, 8 záložníkov, 10 útočníkov. Koľko rôznych jedenástok môže zostaviť, ak potrebuje 1 brankára, 2 obrancov, 4 záložníkov a 4 útočníkov?

Riešenie. $C_1(3) \cdot C_2(5) \cdot C_4(8) \cdot C_4(10) = 3 \cdot 10 \cdot 70 \cdot 210 = 441\,000$.

Úlohy

1.1 Koľkými spôsobmi môžeme navliecť na zvislý drôtik 4 biele, 3 žlté a 2 čierne koráliky? [1260]

1.2 Koľkými spôsobmi môžeme vybrať dvojčlennú nočnú hliadku z 20 skautov? [190]

1.3 Koľkými spôsobmi môžeme zo siedmich chlapcov a štyroch dievčat zostaviť šesťčlenné volejbalové družstvo tak, aby v ňom boli:

- a) práve dve dievčatá? b) aspoň dve dievčatá?

[a)210; b)371]

1.4 V triede je 19 chlapcov a 12 dievčat. Koľkými spôsobmi z nich môžeme vybrať trojčlennú skupinu, v ktorej sú:

- a) iba chlapci? b) iba dievčatá? c) dvaja chlapci a jedno dievča?

[a)969; b)220; c)2052]

1.5 Na policičku chceme postaviť vedľa seba do radu osem slovenských a štyri české knihy. Koľkými spôsobmi môžeme zoradiť:

- a) všetky knihy?
b) tak, aby na ľavom okraji policičky boli vedľa seba české a vpravo slovenské knihy?
c) tak, aby české knihy boli vedľa seba a slovenské tiež?

[a)479 001 600; b)96 768; c)1 935 360]

1.6 V zásielke s päťdesiatimi výrobkami je 6 chybných. Koľkými spôsobmi môžeme vytiahnuť 5 výrobkov tak, aby medzi vytiahnutými boli:

- a) iba dobré výrobky?
- b) 2 dobré a 3 chybné výrobky?
- c) najviac dva chybné výrobky?

[a)1 086 008; b)18 920; c)2 099 174]

1.7 V nádobe je 8 bielych, 10 modrých a 5 červených gúľ. Koľkými spôsobmi môžem z nádoby vybrať:

- a) dve biele gule? b) dve modré gule? c) dve biele a dve modré gule?

[a)28; b)45; c)1 260]

1.8 Jana má 5 rôznych farebných tričiek a tri rôzne sukne. Koľkými spôsobmi sa môže obliecť, aby vždy vyzerala inak? [15]

1.9 V reštaurácii majú na jedálnom lístku 3 druhy polievok, 7 možností výberu hlavného jedla, 4 druhy múčnikov. Na pitie si možno objednať kávu, limonádu alebo džús. Koľkými spôsobmi si hosť môže vybrať obed za predpokladu, že bude jesť:

- a) len polievku a hlavné jedlo,
- b) polievku, hlavné jedlo a ešte si objedná aj nápoj,
- c) polievku, hlavné jedlo, múčnik a nápoj.

[a)21; b)63; c)252]

1.10 Mama kúpila 10 rožkov a 8 žemlí. Martin si vezme buď rožok, alebo žemľu. Potom si Dávid vezme jednu žemľu a jeden rožok. Kedy má Dávid viac možností výberu, ak si Martin vzal rožok, alebo keď si Martin vzal žemľu?

[rožok – 72 možností, žemľa – 70 možností]

1.11 Určte počet všetkých trojciferných prirodzených čísel, ak v ich desiatkovom zápise sa každá číslica vyskytuje najviac raz. [648]

1.12 Z mesta A do mesta B vedú štyri turistické trasy, z mesta B do mesta C vedú tri trasy.

- a) Určte, koľkými spôsobmi môže turista prejsť z A do C, ak sa chce zastaviť v meste B.
- b) Koľkými spôsobmi môže prejsť z A do B a späť, ak nechce ísť späť tou istou cestou?

[a)12; b)12]

1.13 Na vrchol hory vedú 4 turistické cesty a lanovka. Určte počet spôsobov, ktorými sa možno dostať:

- a) na vrchol a späť;
- b) na vrchol a späť tak, aby spiatočná cesta bola iná než cesta na vrchol;
- c) na vrchol a späť tak, aby aspoň raz bola použitá lanovka;
- d) na vrchol a späť tak, aby lanovka bola použitá práve raz.

[a)25; b)20; c)9; d)8]

- 1.14** Koľko 5-miestnych čísel bez opakovania možno zostaviť z číslic: 0, 1, 3, 4, 7? Koľko z nich je párnych? [96; 32]
- 1.15** Koľko 6-miestnych čísel bez opakovania možno zostaviť z číslic: 1, 2, 3, 4, 5, 6, ak sa čísla majú začínať: a) číslicou 4; b) číslicou 4 alebo 5? [a)120; b)240]
- 1.16** Koľko jedno až 4-miestnych čísel možno zostaviť z číslic: 0, 2, 4, 6? [49]
- 1.17** Koľko je všetkých trojciferných prirodzených čísel? [900]
- 1.18** Koľko rôznych signálov možno utvoriť z piatich zástaviek rôznych farieb, ak vedľa seba stoja: a) tri zástavky (trikolóry); b) 2 zástavky (bikolóry)? [a)60; b)20]
- 1.19** Koľkými spôsobmi možno odmeniť 1., 2. a 3. cenou 13 účastníkov športovej súťaže? [1716]
- 1.20** Koľkými spôsobmi môžeme usadiť za stôl 5 hostí? [120]
- 1.21** V lavici sedí 5 žiakov, z ktorých dvaja sú kamaráti. Koľkými spôsobmi ich môžeme posadiť, aby kamaráti sedeli vedľa seba? [48]
- 1.22** V obchode majú 9 druhov pohľadníc. Koľkými spôsobmi možno kúpiť 14? [319 770]
- 1.23** Na poličke treba zostaviť vedľa seba 3 zelené, 2 červené a 2 žlté hrnčeky.
a) Koľko rôznych spôsobov rozostavenia môže vzniknúť?
b) Koľko spôsobov vznikne, ak hrnčeky rovnakej farby stoja vedľa seba? [a)210; b)6]
- 1.24** Koľkými spôsobmi môžete zostaviť 5-členné basketbalové družstvo chlapcov, 5-členné družstvo dievčat a 5-členné družstvo s dvomi chlapcami a tromi dievčatami, ak máte k dispozícii 7 chlapcov a 8 dievčat? [21; 56; 1 176]
- 1.25** Koľkými spôsobmi možno zostaviť družstvo obsahujúce troch chlapcov a tri dievčatá z triedy, v ktorej je 15 chlapcov a 10 dievčat? [54 600]
- 1.26** Osem študentov má pripravené ubytovanie na internáte v troch izbách - dve sú 3-posteľové, jedna 2-posteľová. Koľko je spôsobov rozdelenia študentov do jednotlivých izieb? [560]
- 1.27** V rýchlikovej vlakovej súprave sú dva batožinové vozne, jeden jedálenský vozeň, štyri lôžkové vozne a tri ležadlové vozne. Koľkými spôsobmi možno zoradiť vozne súpravy? [12 600]
- 1.28** V osudí je 35 lístkov označených číslami 1 až 35. Postupne z nich vytiahneme päť, ale lístky do osudia nevraciamy späť. Pritom záleží na poradí vytiahnutých čísel. Určte počet všetkých päťíc čísel, ktoré sa môžu vytiahnuť. [38 955 840]
- 1.29** Vlajka má byť zložená z troch rôznofarebných vodorovných pruhov. K dispozícii máme biele, červené, modré, zelené a žlté pruhy. Určte:
a) počet rôznych vlajok, ktoré môžeme z týchto pruhov zložiť;

b) počet rôznych vlajok s modrým pruhom;

c) počet rôznych vlajok, ktoré majú modrý pruh v prostriedku;

d) počet rôznych vlajok, ktoré nemajú červený pruh v prostriedku.

[a) 60; b) 36; c) 12; d) 48]

1.30 O telefónnom čísle svojho spolužiaka vedel Janko len to, že je šesťmiestne, začína sedmičkou, neobsahuje dve rovnaké číslice a je deliteľné 25. Určte, koľko telefónnych čísel prichádza do úvahy? [420]

1.31 Štátnu poznávaciu značku v istom štáte tvoria tri písmená (z 28 možných) a štyri číslice. Určte, koľko poznávacích značiek je v tomto štáte k dispozícii. [219 520 000]

1.32 Kufrík má zámok na heslo, ktorý sa otvorí len vtedy, ak v každom z piatich kotúčov nastavíme správnu číslicu. Na každom kotúči je 9 číslic. Určte najväčší možný počet pokusov. Ktoré musíme vykonať, ak chceme kufrík otvoriť, keď sme zabudli heslo. [59 049]

1.33 Na štarte bežeckého preteku je 8 atlétov. Koľkými spôsobmi môžu byť obsadené stupne víťazov? [336]

2. TEÓRIA PRAVDEPODOBNOTI

2.1 NÁHODNÝ POKUS, NÁHODNÝ JAV

Teória pravdepodobnosti sa zaoberá matematickými zákonitosťami, ktoré sa prejavujú v náhodných pokusoch. Tieto zákonitosti majú opodstatnenie iba pri dostatočne veľkom počte pokusov.

Náhodné pokusy sú pokusy, ktoré pri dodržaní určitých vopred daných podmienok privedú k rôznym výsledkom. Tieto výsledky závisia nielen od predpísaných podmienok, ktoré sa pri každej realizácii pokusu nemenia, ale aj od náhody.

Náhodný pokus teda musí mať vlastnosť hromadnosti, musí sa dať opakovať nekonečne veľakrát pri dodržaní rovnakých podmienok.

Pre každý náhodný pokus je možné dopredu vymenovať všetky možné výsledky, a to tak, že sa navzájom vylučujú a jeden z nich nastane vždy. Táto množina najjednoduchších, najelementárnejších, navzájom rôznych, základných možných výsledkov sa označuje Ω . Ľubovoľný prvok množiny Ω sa nazýva elementárny jav a označuje sa $E_1, E_2, \dots$ (niekedy sa používa označenie $\omega_1, \omega_2, \dots$). Množina všetkých elementárnych javov, čiže $\Omega = \{E_1, E_2, \dots, E_n, \dots\}$ sa nazýva priestor elementárnych javov. Ľubovoľná podmnožina Ω sa nazýva náhodný jav. Náhodné javy sa označujú veľkými písmenami zo začiatku abecedy $A, B, C, \dots, A_1, A_2, \dots$. Prázdna podmnožina $\emptyset$ množiny Ω sa nazýva nemožný jav, t.j. jav, ktorý nikdy nenastane. Celá množina (ako podmnožina) Ω je istý jav (označuje sa I), t.j. jav, ktorý nastane pri každom náhodnom pokuse.

Pre takto definované javy platí všetko, čo platí pre množiny. Môžeme definovať operácie medzi javmi, môžeme graficky znázorňovať javy, atď., t.j. nájdeme interpretáciu základných pojmov a vzťahov z teórie množín v rámci teórie náhodných javov.

V ďalšom pokiaľ budeme hovoriť o dvoch, resp. viacerých náhodných javoch, budeme uvažovať vždy o javoch z toho istého základného priestoru elementárnych javov.

2.2 VZŤAHY MEDZI JAVMI

Nech A, B sú javy (podmnožiny Ω).

Jav A je podjavom (časťou) javu B , ak z nastania javu A vyplýva nastanie javu B , čo zapisujeme $A \subset B$.

Javy A a B sa rovnajú, t.j. $A = B$, ak $A \subset B \wedge B \subset A$.

Jav $\bar{A}$ je opačný jav k javu A , ak jav $\bar{A}$ nastane práve vtedy, keď nenastane jav A .

Jav $C = A \cap B$ sa nazýva prienik alebo súčin javov A a B . Jav C nastane práve vtedy, keď nastane jav A a súčasne jav B .

Jav $C = A \cup B$ sa nazýva zjednotenie alebo súčet javov A a B . Jav C nastane práve vtedy, keď nastane jav A alebo B .

Javy A a B sú disjunktné (nezlučiteľné), ak $A \cap B = \emptyset$, t.j. ich súčasné nastanie nie je možné.

Skupina javov $A_1, A_2, \dots, A_n$ tvorí úplný systém javov, ak pre ľubovoľnú dvojicu platí:

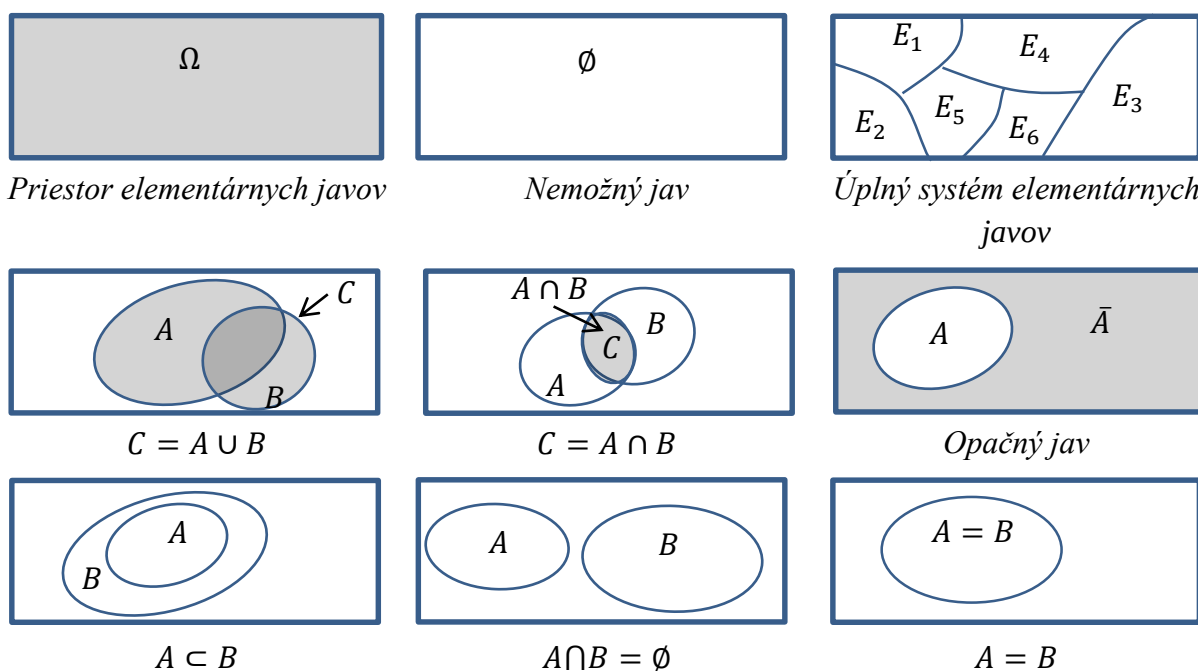
1. $A_i \cap A_j = \emptyset; i, j = 1, 2, \dots, n \quad i \neq j$
2. $A_1 \cup A_2 \dots \cup A_n = I$

Jav C sa nazýva zložený jav, ak existujú aspoň 2 javy A, B rôzne od nemožného javu tak, že $C = A \cup B$ pričom $A \neq C; B \neq C$.

Na základe vyššie uvedených definícií môžeme vysloviť nasledujúce vlastnosti elementárnych javov:

1. Jav $E \neq \emptyset$ je elementárny jav práve vtedy, keď neexistuje taký jav A , pre ktorý by platilo $A \subset E$.
2. Rôzne elementárne javy sú navzájom disjunktné.
3. Zjednotenie všetkých elementárnych javov je jav istý (čiže elementárne javy tvoria úplný systém javov).
4. Každý zložený jav sa dá jednoznačne vyjadriť ako zjednotenie aspoň dvoch elementárnych javov.

Grafické zobrazenie javov a jednotlivých operácií s javmi pomocou Vennovho diagramu :



Obr. 2.1 Znáozornenie náhodných javov pomocou Vennových diagramov

Príklad 2.1. Nech náhodný pokus je hod hracou kockou.

- a) Zostrojte priestor elementárnych javov.
- b) Nech sú dané javy A – padne párny počet bodov; B – padne 2, 4 alebo 5; C – padne počet bodov 7; D – padne prirodzené číslo. Vyjadrite dané javy pomocou elementárnych javov. Nájdite opačný jav k javu A , súčin a súčet javov A, B .

Riešenie. a) Označme elementárne javy, čiže navzájom rôzne možné výsledky pokusu E_1 – padne počet bodov 1, E_2 – padne počet bodov 2, E_3 – padne počet bodov 3, ..., E_6 – padne počet bodov 6. Potom priestor elementárnych javov je $\Omega = \{E_1, E_2, E_3, E_4, E_5, E_6\}$.

b) jav $A = E_2 \cup E_4 \cup E_6$,
jav $B = E_2 \cup E_4 \cup E_5$, čo niekedy zapisujeme aj krátko $B = (E_2, E_4, E_5)$,
súčet javov $A \cup B = \{E_2, E_4, E_5, E_6\}$,
súčin javov $A \cap B = \{E_2, E_4\}$,
jav $\bar{A}$ – opačný jav k javu A znamená, že padne nepárny počet bodov, t.j. $\bar{A} = E_1 \cup E_3 \cup E_5$,
jav $C = \emptyset$ je nemožný jav,
jav $D = I$ je istý jav.

2.3 PRAVDEPODOBNOŠŤ A JEJ VLASTNOSTI

Definícia 2.1 Nech Ω je priestor elementárnych javov $\Omega \neq \emptyset$. Nech S je množina podmnožín Ω s vlastnosťami:

1. $\Omega \in S$
2. $A \in S, B \in S \Rightarrow A \cup B \in S, A \cap B \in S, \bar{A} \in S, \bar{B} \in S$

Potom S sa nazýva javové pole.

Definícia 2.2 (Kolmogorova axiomatická definícia pravdepodobnosti)

Nech je daný priestor elementárnych javov $\Omega \neq \emptyset$ javové pole S . Pravdepodobnosťou rozumieme ľubovoľnú reálnu funkciu $P: S \rightarrow R$, ktorá má vlastnosti:

- Axióma 1: Každému javu $A \in S$ je priradené nezáporné číslo $P(A) \geq 0$.
- Axióma 2: Pravdepodobnosť istého javu je rovná jednej; $P(I) = 1$.
- Axióma 3: Ak $A_1, A_2, \dots, A_n, \dots$ je postupnosť disjunktných javov S , t.j. $A_i \cap A_j = \emptyset$ pre $\forall i \neq j$, potom $P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$.

Trojica (Ω, S, P) sa nazýva pravdepodobnostný priestor.

Veta 2.1 (Vlastnosti pravdepodobnosti)

Nech je daný pravdepodobnostný priestor (Ω, S, P) a nech $A, B \in S$. Potom platí:

1. $P(\emptyset) = 0$
2. $P(\bar{A}) = 1 - P(A)$
3. $A \subset B \Rightarrow P(A) \leq P(B)$
4. $0 \leq P(A) \leq 1$

Definícia 2.3 (Klasická definícia pravdepodobnosti)

Predpokladajme, že $\Omega = \{E_1, E_2, \dots, E_n\}, n \in N$ je konečná množina a všetky elementárne javy sú rovnako možné. Nech A je jav, ktorý sa dá rozložiť na k elementárnych javov. Potom pravdepodobnosť javu A je

$$P(A) = \frac{k}{n} = \frac{\text{počet elementárnych javov priaznivých javu } A}{\text{počet všetkých možných elementárnych javov}}.$$

Poznámka 2.1 Dá sa dokázať, že takto definovaná funkcia pravdepodobnosti vyhovuje všetkým trom axiómam Kolmogorovej definície pravdepodobnosti.

Príklad 2.2 Hodíme hracou kockou. Nech jav A znamená, že padne počet bodov menší ako 3. Aká je pravdepodobnosť nastania javu A ?

Riešenie. Priestor elementárnych javov bude $\Omega = \{E_1, E_2, \dots, E_6\}$, ktorá je konečná, $n = 6$, kde E_i znamená, že na kocke padne počet bodov $i, i = 1, 2, \dots, 6$. Jav $A = \{E_1 \cup E_2\}$ sa dá rozložiť na $k = 2$ elementárne javy. Podľa klasickej definície pravdepodobnosť javu A je číslo $P(A) = \frac{2}{6} = \frac{1}{3}$.

Definícia 2.4 (Podmienená pravdepodobnosť)

Nech je daný pravdepodobnostný priestor (Ω, S, P) a nech $A, B \in S, P(B) > 0$. Potom podmienená pravdepodobnosť javu A/B , t.j. javu A pri podmienke, že jav B nastal, sa rovná

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

Poznámka 2.2 Takto definovaná podmienená pravdepodobnosť vyhovuje axiomatickej definícii pravdepodobnosti, preto má všetky vlastnosti pravdepodobnosti.

Definícia 2.5 Nech je daný pravdepodobnostný priestor (Ω, S, P) . Nech $A, B \in S$ sú javy, pre ktoré $P(A) > 0, P(B) > 0$. Hovoríme, že javy A a B sú nezávislé, ak $P(A/B) = P(A)$ a $P(B/A) = P(B)$.

Veta 2.2 Nech je daný pravdepodobnostný priestor (Ω, S, P) . Nech $A, B \in S$. Potom platí

1. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (veta o sčítaní pravdepodobností),
2. $P(A \cap B) = P(A/B) \cdot P(B) = P(B/A) \cdot P(A)$ (veta o násobení pravdepodobností),
3. Javy A, B sú vzájomne nezávislé práve vtedy, keď $P(A \cap B) = P(A) \cdot P(B)$ (kritérium nezávislosti).

Príklad 2.3 Hádzame trikrát 2- eurovou mincou. Nech jav A znamená, že padne aspoň dvakrát znak. Aká je pravdepodobnosť javu A ?

Riešenie. Znázorníme strom logických možností daného pokusu.



Nech Z znamená, že padne znak, H znamená, že padne hlava, potom priestor elementárnych javov $\Omega = \{(Z, Z, Z), (H, Z, Z), (Z, H, Z), (Z, Z, H), (H, H, Z), (H, Z, H), (Z, H, H), (H, H, H)\}$ je konečný, $n = 8$.

Elementárne javy, z ktorých sa skladá jav A sú $A = \{(Z, Z, Z), (H, Z, Z), (Z, H, Z), (Z, Z, H)\}$, ich počet je $k = 4$. Podľa klasickej definície pravdepodobnosti dostaneme pravdepodobnosť javu A číslo $P(A) = \frac{4}{8} = 0,5$, vyjadrené v percentách 50%.

Poznámka 2.3 Ak označíme jednotlivé elementárne javy A_1 až A_8 nasledovne:

$\Omega = \{A_1 = (Z, Z, Z), A_2 = (H, Z, Z), A_3 = (Z, H, Z), A_4 = (Z, Z, H), A_5 = (H, H, Z), A_6 = (H, Z, H), A_7 = (Z, H, H), A_8 = (H, H, H)\}$, potom jav $A = A_1 \cup A_2 \cup A_3 \cup A_4$.

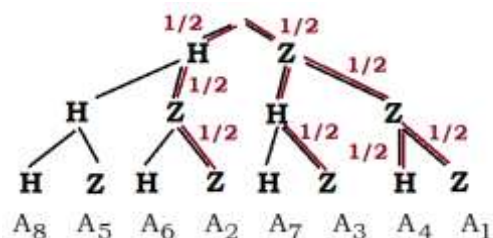
Elementárne javy sú disjunktné, preto platí 3. axióma pravdepodobnosti, podľa ktorej $P(A) = P(A_1 \cup A_2 \cup A_3 \cup A_4) = P(A_1) + P(A_2) + P(A_3) + P(A_4)$.

Z hore uvedeného označenia vyplýva, že jav A_1 označuje výsledok keď trikrát padne znak Z $A_1 = (Z, Z, Z)$, t.j. značí súčasné nastanie javu B_1 – prvý raz padne znak Z , javu B_2 – druhý raz padne znak Z a javu B_3 – tretí raz padne znak Z . Jav $A_1 = B_1 \cap B_2 \cap B_3$, jav A_1 je súčinom troch nezávislých javov, preto podľa vety o násobení pravdepodobnosti dostaneme

$$P(A_1) = P(B_1 \cap B_2 \cap B_3) = P(B_1) \cdot P(B_2) \cdot P(B_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

Analogicky možno určiť pravdepodobnosti ostatných elementárnych javov $P(A_1) + P(A_2) + P(A_3) + P(A_4) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{4}{8} = 0,5$.

Ak doplníme do stromu logických možností prislúchajúce pravdepodobnosti, dostaneme pravdepodobnostný strom.



Príklad 2.4 Zväzok kľúčov obsahuje 5 podobných kľúčov, z ktorých iba jeden otvára určitú zámku. Aká je pravdepodobnosť, že sa zámka podarí otvoriť:

- pri náhodnom výbere jedného kľúča na prvý pokus,
- pri náhodnom výbere dvoch kľúčov až na druhý pokus.

Riešenie. a) Nech jav A znamená, že zámku sme otvorili na prvý pokus prvým náhodne vybraným kľúčom. Podľa klasickej definície pravdepodobnosti $P(A) = \frac{1}{5} = 0,2$.

b) Nech jav A znamená, že zámku sme neotvorili na prvý pokus prvým náhodne vybraným kľúčom. Nech jav B znamená, že zámku sme otvorili na druhý pokus druhým náhodne vybraným kľúčom. Hľadáme pravdepodobnosť javu $A \cap B$, t.j. prvým kľúčom sme zámku neotvorili a druhým kľúčom sme už otvorili. Podľa podmienenej pravdepodobnosti bude $P(A \cap B) = P(A) \cdot P(B/A) = \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{5} = 0,2$.

2.4 OPAKOVANÉ NÁHODNÉ POKUSY

Rozoznávame dva druhy opakovaných náhodných pokusov:

- Opakované nezávislé pokusy, pri ktorých výsledok každého jednotlivého pokusu nezávisí od výsledku predchádzajúcich pokusov. Príkladom takýchto pokusov je výber s opakovaním.
- Opakované závislé pokusy, pri ktorých výsledok každého jednotlivého pokusu závisí od výsledku predchádzajúcich pokusov. Príkladom takýchto pokusov je výber bez opakovania.

2.4.1 Nezávislé opakované pokusy

Veta 2.3 (Bernoulliho schéma)

Nech pravdepodobnosť nastania javu A je v každom pokuse rovnaká a rovná sa pravdepodobnosti $P(A) = p$. Potom pravdepodobnosť $P_{k,n}$ toho, že pri n pokusoch jav A nastane práve k – krát ($0 \leq k \leq n$) sa rovná $P_{k,n} = \binom{n}{k} p^k (1-p)^{n-k}$.

Pre výpočet $P_{k,n}$ slúži v Exceli funkcia = BINOM.DIST($k; n; p$; kumulatívne), kde argumentu kumulatívne priradíme logickú hodnotu FALSE.

Všimnime si, že vzorec pre $P_{k,n}$ je $(k+1)$ -vý člen binomického rozvoja $(p + (1-p))^n$, odtiaľ názov funkcie v Exceli (pozri binomické rozdelenie).

Poznámka 2.4 Ak počítame pravdepodobnosť toho, že pri n pokusoch jav A nastane najviac k – krát, potom $P(A) = P_{0,n} + P_{1,n} + \dots + P_{k,n} = \sum_{i=1}^k P_{i,n}$.

V Exceli hodnota $P(A) = \text{BINOM.DIST}(k; n; p)$, kde argumentu kumulatívne priradíme logickú hodnotu TRUE.

Ak počítame pravdepodobnosť toho, že pri n pokusoch jav A nastane aspoň k -krát, pravdepodobnosť javu dostaneme ako

$$P(A) = P_{k,n} + P_{k+1,n} + \dots + P_{n,n} = 1 - (P_{0,n} + P_{1,n} + \dots + P_{k-1,n}) = 1 - \sum_{i=1}^{k-1} P_{i,n}.$$

V Exceli hodnotu $P(A)$ vypočítame pomocou funkcie = 1 – BINOM.DIST($k - 1; n; p$), kde argumentu kumulatívne priradíme logickú hodnotu TRUE.

Príklad 2.5 Športový strelec zasiahne cieľ (pri každom výstrele) s pravdepodobnosťou 0,85.

Vypočítajte pravdepodobnosť, že pri ôsmich výstreloch bude mať v cieľi:

- a) práve päť zásahov, b) aspoň dva zásahy, c) najviac 3 zásahy.

Riešenie. Pravdepodobnosť, že pri ôsmich výstreloch bude v cieľi:

a) práve päť zásahov je

$$P_{5,8} = \binom{8}{5} 0,85^5 (1 - 0,85)^{8-5} = \binom{8}{5} \cdot 0,85^5 \cdot 0,15^3 = \text{BINOM.DIST}(5; 8; 0,85; 0) =$$

0,08386;

b) aspoň dva zásahy

$$P_{2,8} + P_{3,8} + \dots + P_{8,8} = 1 - (P_{2,8} + P_{2,8}) = 1 - \binom{8}{0} 0,85^0 (1 - 0,85)^{8-0} - \binom{8}{1} 0,85^1 (1 - 0,85)^{8-1} = 1 - 0,15^8 - 8 \cdot 0,85 \cdot 0,15^7 = 1 - \text{BINOM.DIST}(1; 8; 0,85; 1) = 0,999988 ;$$

c) najviac 3 zásahy

$$P_{0,8} + P_{1,8} + P_{2,8} + P_{3,8} = \binom{8}{0} 0,85^0 (1 - 0,85)^{8-0} + \binom{8}{1} 0,85^1 (1 - 0,85)^{8-1} + \binom{8}{2} 0,85^2 (1 - 0,85)^{8-2} + \binom{8}{3} 0,85^3 (1 - 0,85)^{8-3} = \text{BINOM.DIST}(3; 8; 0,85; 1) = 0,002853 .$$

Príklad 2.6 Prístroj je vyradený z prevádzky, keď sa prepáli aspoň 5 elektrónok prvého typu alebo aspoň 2 elektrónky druhého typu. Určte pravdepodobnosť, že prístroj bude vyradený z prevádzky, keď vieme, že sa prepáli 5 elektrónok a že pravdepodobnosť prepálenia elektrónok prvého typu je 0,7 a druhého typu je 0,3.

Riešenie. Nech jav A znamená, že v prístroji sa prepáli 5 elektrónok prvého typu. Podľa Bernoulliho schémy dostaneme

$$P(A) = \binom{5}{5} 0,7^5 (1 - 0,7)^0 = 0,7^5 = \text{BINOM.DIST}(5; 5; 0,7; 0) = 0,16807.$$

Nech jav B označuje, že sa prepália aspoň 2 (čiže 2 alebo 3, alebo 4, alebo 5) elektrónok druhého typu. Na základe Bernoulliho schémy pravdepodobnosť javu B je

$$P(B) = \binom{5}{2} 0,3^2 (1 - 0,3)^3 + \binom{5}{3} 0,3^3 (1 - 0,3)^2 + \binom{5}{4} 0,3^4 (1 - 0,3)^1 + \binom{5}{5} 0,3^5 (1 - 0,3)^0 = 1 - \text{BINOM.DIST}(1; 5; 0,3; 1) = 0,47175 .$$

Javy A a B sú disjunktné, hľadáme pravdepodobnosť nastania javu A alebo B , čo podľa vety o pravdepodobnosti súčtu je

$$P(A \cup B) = P(A) + P(B) = 0,16807 + 0,47175 = 0,63982 .$$

2.4.2 Závislé opakované pokusy

Veta 2.4 (*Hypergeometrická formula*)

Daných je N predmetov, z ktorých M ($0 < M < N$) predmetov má určitý znak (vlastnosť). Z celého množstva N predmetov sa náhodne vyberie n ($0 < n < N$) predmetov. Nech jav A znamená, že v tomto výbere n predmetov má práve k predmetov uvažovaný znak. Potom pravdepodobnosť javu A je určená hypergeometrickou formulou

$$P(A) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} .$$

Pre výpočet pravdepodobnosti nastania javu A , t.j. $P(A)$ je v Exceli preddefinovaná funkcia = HYPGEOM.DIST (vzorka_s; veľkosť_vzorky; populácia_s; počet_pop; kumulatívne), kde parametru vzorka_s priradíme hodnotu k , veľkosť_vzorky je rozsah výberu n , populácia_s je M , počet_pop je celkové množstvo N , argumentu kumulatívne priradíme logickú hodnotu TRUE.

Priklad 2.7 V nádobe sa nachádza 10 ložísk, z nich 6 je guľôčkových, ostatné sú valčekové. Náhodne vyberieme z nádoby bez opakovania 3 ložiská. Aká je pravdepodobnosť, že z nich dve budú valčekové?

Riešenie. V našom prípade počet všetkých ložísk je $N = 10$, z nich valčekové sú $M = 4$. Náhodne vyberieme $n = 3$ ložiská a z nich $k = 2$ budú valčekové. Podľa hypergeometrickej formuly dostaneme pravdepodobnosť javu A – dve ložiská z 3 vybraných sú valčekové

$$P(A) = \frac{\binom{4}{2} \cdot \binom{10-4}{3-2}}{\binom{10}{3}} = \text{HYPGEOM.DIST}(2; 3; 4; 10; 0) = 0,3.$$

Úlohy

2.1 Aká je pravdepodobnosť, že pri hode mincou padne znak? $[0,5]$

2.2 Aká je pravdepodobnosť, že pri hode očíslovaným štvorstenom (číslami 1 až 4) padne číslo 1? $[0,25]$

2.3 Aká je pravdepodobnosť, že pri hode hracou kockou (čísla 1 až 6) padne číslo väčšie než 4, menšie než 4 a rovné 4? $\left[\frac{1}{3}; 0,5; \frac{1}{6}\right]$

2.4 Hracia kocka typu $\{1, 1, 1, 2, 2, 3\}$ je kocka, ktorá má na troch stranách 1, na dvoch stranách 2 a na jednej strane 3. Predpokladajme len 6 možností jej pádu. Aká je pravdepodobnosť, že pri hode touto hracou kockou padne číslo 1 a aká je pravdepodobnosť, že padne číslo 2? $\left[0,5; \frac{1}{3}\right]$

2.5 Na policike je náhodne rozostavených 10 kníh. Aká je pravdepodobnosť, že určité tri knihy sú postavené vedľa seba? $[0,066]$

2.6 Chlapec sa hrá s písmenami A, A, E, I, K, A, T, M, M, T. Aká je pravdepodobnosť, že pri náhodnom usporiadaní písmen dostane slovo MATEMATIKA? $[6,61 \cdot 10^{-6}]$

2.7 Čísllice 1, 2, 3, 4, 5 sú napísané na piatich lístkoch. Náhodne vyberieme tri lístky a uložíme vedľa seba v poradí, v akom sme ich vybrali. Aká je pravdepodobnosť toho, že takto vzniknuté trojciferné číslo bude párne? $[0,4]$

2.8 V ukradnutej peňaženke bola aj bankomatová karta. S akou pravdepodobnosťou zlodej na prvý pokus uhádne jej štvormiestny kód PIN? $[0,0001]$

2.9 Aká je pravdepodobnosť, že ľubovoľné číslo sa končí:
a) päťkou; b) nulou; c) párnou číslicou? $[0,1; 0,1; 0,5]$

2.10 V meste sú štyri križovatky so svetelnými semaformi. Každý z nich uvoľňuje alebo uzatvára dopravu s rovnakou pravdepodobnosťou 0,5. Aká je pravdepodobnosť, že auto:

- a) prejde prvou križovatkou bez zdržania;
- b) prejde prvými dvoma križovatkami bez zdržania;
- c) prejde všetkými štyrmi križovatkami bez zdržania? [0,5; 0,25; 0,0625]

2.11 V osudí sú 4 biele a 3 modré lístky. Náhodne vyberieme dva lístky. Aká je pravdepodobnosť, že budú:

- a) obidva biele ; b) obidva modré; c) jeden biely a jeden modrý? $\left[\frac{2}{7}; \frac{1}{7}; \frac{4}{7}\right]$

2.12 V zásielke 20 rozhlasových prijímačov je 10% chybných. Vypočítajte pravdepodobnosť, že pri predaji 3 prijímačov sú 2 bez chýb. [0,269]

2.13 Študent si pri skúške losuje 2 z desiatich otázok. Pripravený je na 6 z nich. Aká je pravdepodobnosť, že:

- a) bude vedieť obidve;
- b) bude vedieť aspoň jednu z vylosovaných otázok;
- c) bude vedieť práve jednu;
- d) nebude vedieť žiadnu? $\left[\frac{1}{3}; \frac{1}{15}; \frac{2}{15}; \frac{2}{5}\right]$

2.14 Každá vzorka vody má 10% pravdepodobnosti toho, že obsahuje organickú znečisťujúcu látku. Predpokladáme, že vzorky sú nezávislé vzhľadom k prítomnosti znečisťujúcej látky. Nájdite pravdepodobnosť, že z 18 vzoriek:

- a) práve 2 budú obsahovať znečisťujúcu látku,
- b) aspoň 4 vzorky budú obsahovať znečisťujúcu látku. [a) 0,284; b) 0,098]

2.15 Mladý otec by chcel mať syna. Naplánoval si 3 deti. Z dlhodobých štatistík je známe, že pravdepodobnosť narodenia chlapca je 0,51. Aká je pravdepodobnosť, že z jeho 3 detí bude mať práve: a) 1 syna; b) 2 synov; c) všetkých 3 synov? [a) 0,367; b) 0,382, c) 0,133]

2.16 Pravdepodobnosť narodenia chlapca je 0,515. Určte počet detí, aby medzi nimi bol aspoň jeden chlapec s pravdepodobnosťou rovnajúcou sa aspoň 0,99. [7]

2.17 Pri sedemnásobnom nezávisle opakovanom hode hracou kockou vypočítajte pravdepodobnosť toho, že šestka padne: a) práve dvakrát, b) aspoň dvakrát, c) najviac dvakrát. [a) 0,2344; b) 0,33; c) 0,903]

2.18 Písomná skúška z matematiky má päť príkladov. Pravdepodobnosť správneho vypočítania jedného príkladu je 0,8. Aká je pravdepodobnosť, že študent na skúške vyhoví, ak na to aby vyhovel, stačí vypočítať tri príklady? [0,942]

2.19 Písomku píše 12 študentov. Pravdepodobnosť, že v ľubovoľnom okamihu pracuje študent sústredene je 0,8. Aká je pravdepodobnosť, že v danom okamihu pracuje sústredene aspoň 10 študentov? [0,5584]

2.20 Klíčivosť semien určitého druhu je 0,98. K pokusným účelom je zasiatych 10 semien. Aká je pravdepodobnosť, že z nich vyklíči 8 semien? [0,0153]

- 2.21 Liek úspešne lieči ochorenie v 80% prípadoch. Liek sa podáva 10 pacientom. Aká je pravdepodobnosť, že sa aspoň 7 z nich vylieči? [0,8791]
- 2.22 Pravdepodobnosť toho, že pri meraní danej fyzikálnej veličiny presiahne chyba dovolenú odchýlku je 0,4. Urobia sa 3 nezávislé merania. Aká je pravdepodobnosť, že chybu, ktorá presiahne povolenú odchýlku, urobíme práve pri jednom meraní? [0,432]
- 2.23 Zle zostavený automat vyrába 30% nepodarkov. Aká je pravdepodobnosť, že pri kontrole troch výrobkov: a) nenájde žiadny nepodarok, b) nájdeme 2 nepodarky? [a)34,3%; b)18,9%]
- 2.24 Podnik vyrába 70% produkcie prvej triedy kvality. Aká je pravdepodobnosť, že v sérii 100 výrobkov počet výrobkov prvotriednej kvality je od 60 do 62? [4%]
- 2.25 Študent má vypracovať test, ktorý obsahuje 10 otázok a ku každej z nich sú 4 odpovede, pričom práve jedna je správna. Aká je pravdepodobnosť, že študent, ktorý látku vôbec nepozná a robí odpovede náhodne, zodpovie správne aspoň 5 otázok? [0,0781]

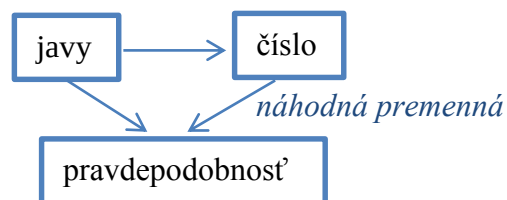
3. NÁHODNÁ PREMENNÁ

V pravdepodobnostnom priestore (Ω, S, P) môže byť každý elementárny jav číselne hodnotený. Toto je možné opísať funkciou $X: \Omega \rightarrow \mathbb{R}$.

Príklad 3.1 Určte, aká bude výška mojej výhry pri hádzaní hracou kockou pri dodržaní nasledujúcich pravidiel:

- ak hodím počet bodov 1 alebo 2 vyhrávam 2 €;
- ak hodím počet bodov 3,4,5 vyhrávam 5 €;
- ak hodím počet bodov 6, prehrávam 10 €.

Riešenie. Priestor elementárnych javov je $\Omega = \{E_1, E_2, E_3, E_4, E_5, E_6\}$, kde elementárny jav E_i znamená padnutie počtu bodov $i, i = 1, 2, \dots, 6$. Pravdepodobnosť každého elementárneho javu je $P(E_i) = \frac{1}{6}$. Nech jav $A = \{E_1, E_2\}$, $B = \{E_3, E_4, E_5\}$, $C = \{E_6\}$. Potom ich pravdepodobnosti budú $P(A) = \frac{2}{6}$, $P(B) = \frac{3}{6} = \frac{1}{2}$, $P(C) = \frac{1}{6}$. Podľa zadania javu A priradíme číslo 2, javu B číslo 5 a javu C záporné číslo -10. Tým sme číselne ohodnotili každý jav. Javy sú náhodné, teda aj čísla im priradené sú náhodné. Ak priradíme týmto číslam zodpovedajúce pravdepodobnosti, vznikne tzv. náhodná premenná. Čiže vznikne zobrazenie z množiny elementárnych javov do množiny reálnych čísel, ktoré môžeme znázorniť ako



Výšku výhry môžeme chápať ako funkciu $X: \Omega \rightarrow \mathbb{R}$ definovanú nasledovne $X(1) = X(2) = 2$, $X(3) = X(4) = X(5) = 5$, $X(6) = -10$, čo môžeme znázorniť v tabuľke

x_i (výška výhry)	2	3	-10	súčet
p_i (pravdepodobnosť výhry)	$\frac{2}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	1

Definícia 3.1 Ak je daný pravdepodobnostný priestor (Ω, S, P) , kde Ω je konečná alebo spočítateľná (tvorí postupnosť), potom náhodná premenná (náhodná veličina) X je ľubovoľná reálna funkcia $X: \Omega \rightarrow \mathbb{R}$. Ak je Ω nespočítateľná, potom zobrazenie X musí byť také, že pre $\forall x \in \mathbb{R}: \{E \in \Omega: X(E) < x\} \in S$.

Rozlišujeme dva typy náhodnej premennej:

- diskretná náhodná premenná – je náhodná premenná, ktorá nadobúda hodnoty, ktoré tvoria konečnú alebo spočítateľnú množinu čísel;
- spojitá náhodná premenná – je náhodná premenná, ktorá nadobúda ľubovoľnú hodnotu z číselného intervalu.

Náhodná premenná bude úplne opísaná, ak poznáme obor hodnôt, (t.j. všetky možné hodnoty, ktoré náhodná premenná nadobúda) a pravdepodobnosti, s ktorými tieto hodnoty nadobúda. (V prípade, že Ω je nespočítateľná, musíme poznať pravdepodobnosť javov $\{E \in \Omega: X(E) < x\}$.) Ak poznáme tieto informácie, hovoríme, že je daný **zákon rozdelenia pravdepodobnosti náhodnej premennej**.

3.1 ZÁKON ROZDELENIA PRAVDEPODOBNOСТИ DISKRÉTNEJ NÁHODNEJ PREMENNEJ

Nech X je diskretná náhodná premenná, jej obor hodnôt je množina $M = \{x_1, x_2, \dots, x_n, \dots\}$. Uvažujeme javy: A_1 – náhodná premenná X nadobúda hodnotu x_1 , t.j. jav $A_1 = (X = x_1)$, jav $A_2 = (X = x_2)$, jav $A_3 = (X = x_3)$, ... $A_n = (X = x_n)$, ...

Javy $A_1, A_2, A_3, \dots, A_n, \dots$ sú disjunktné, t.j. $A_i \cap A_j = \emptyset$ pre $\forall i \neq j$. Označme $P(A_1) = p_1$, $P(A_2) = p_2$, $P(A_3) = p_3$, ..., $P(A_n) = p_n$,

Podľa tretej axiomy pravdepodobnosti pravdepodobnosť zjednotenia týchto javov bude

$$P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$$

Zjednotenie týchto javov je istý jav, preto $P(I) = p_1 + p_2 + \dots + p_n + \dots$, čiže

$$P(I) = \sum_{i=1}^{\infty} p_i = 1.$$

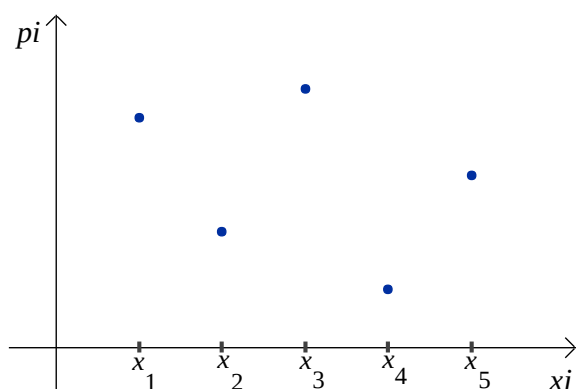
Zákon rozdelenia pravdepodobnosti môže byť zadáný:

1. tabuľkou – radom rozdelenia pravdepodobnosti

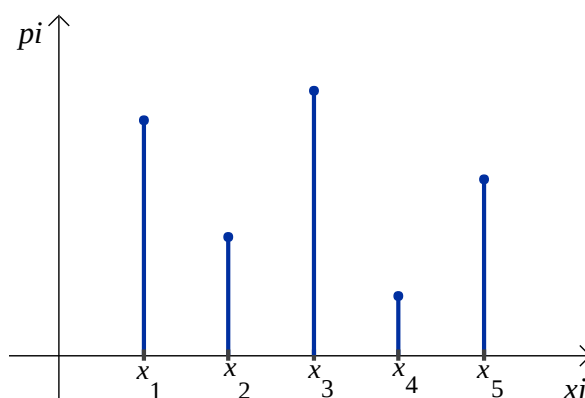
x_i	x_1	x_2	x_3		x_n		Σ
p_i	p_1	p_2	p_3		p_n		1

2. matematickou formulou, napríklad $p(x) = \begin{cases} c \cdot \frac{x}{2} & \text{pre } x = 2, 4, 6, 8 \\ 0 & \text{inak} \end{cases}$

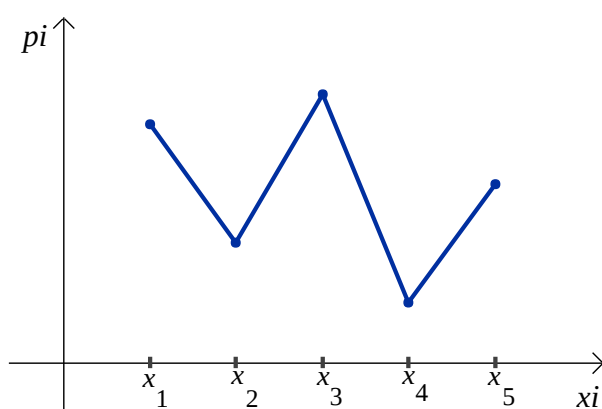
3. graficky



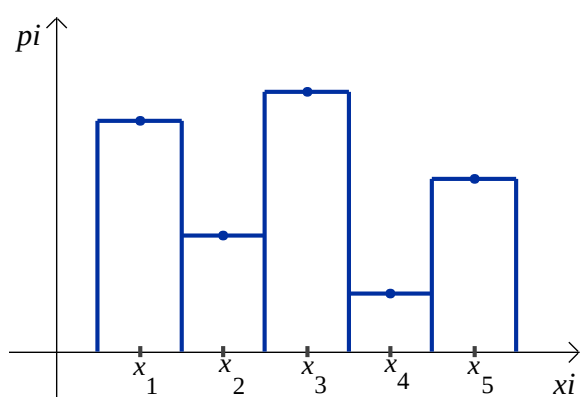
bodovým grafom



paličkovým grafom



polygónom



histogramom

Príklad 3.2 Pre akú hodnotu c sú nasledujúce funkcie pravdepodobnostnými funkciami diskrétnej náhodnej premennej X ?

$$a) p(x) = \begin{cases} c \cdot \left(\frac{1}{4}\right)^x & x = 0, 1, 2, 3, 4 \\ 0 & x \neq 0, 1, 2, 3, 4 \end{cases}$$

$$b) p(x) = \begin{cases} c \cdot \left(\frac{1}{4}\right)^x & x = 0, 1, 2, \dots \\ 0 & x \neq 0, 1, 2, \dots \end{cases}$$

Riešenie. Aby funkcia $p(x) \geq 0$ vyjadrovala zákon rozdelenia pravdepodobnosti náhodnej premennej X formou matematickej formuly, musí platiť, že súčet všetkých pravdepodobností je rovný jednej (pravdepodobnosti istého javu). Na základe toho:

$$\begin{aligned} a) 1 &= \sum_{x=0}^4 c \cdot \left(\frac{1}{4}\right)^x = c \cdot \sum_{x=0}^4 \left(\frac{1}{4}\right)^x = c \cdot \left(\left(\frac{1}{4}\right)^0 + \left(\frac{1}{4}\right)^1 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \left(\frac{1}{4}\right)^4 \right) = \\ &= c \cdot \left(1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{4^4} \right) = c \cdot \left(\frac{4^4 + 4^3 + 4^2 + 4 + 1}{4^4} \right) = c \cdot \frac{341}{256} \Rightarrow c = \frac{256}{341} \end{aligned}$$

$$b) 1 = \sum_{x=0}^{\infty} c \cdot \left(\frac{1}{4}\right)^x = c \cdot \sum_{x=0}^{\infty} \left(\frac{1}{4}\right)^x = c \cdot \left(\left(\frac{1}{4}\right)^0 + \left(\frac{1}{4}\right)^1 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \left(\frac{1}{4}\right)^4 + \dots \right) =$$

$$= c \cdot 1 \cdot \frac{1}{1 - \frac{1}{4}} = c \cdot \frac{1}{\frac{3}{4}} = c \cdot \frac{4}{3} \Rightarrow c = \frac{3}{4}.$$

Výraz v zátvorke $1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{4^4} + \dots$ možno vyjadriť ako súčet nekonečnej geometrickej postupnosti $1, \frac{1}{4}, \frac{1}{4^2}, \frac{1}{4^3}, \frac{1}{4^4}, \dots$, ktorý je určený vzťahom $s = a_1 \cdot \frac{1}{1-q}$, kde a_1 je prvý člen postupnosti $a_1 = 1$ a kvocient $q = \frac{1}{4}$, preto $s = 1 \cdot \frac{1}{1 - \frac{1}{4}}$.

V zadaní a) náhodná premenná X môže nadobudnúť 5 rôznych hodnôt 0, 1, 2, 3, 4, každú s pravdepodobnosťou $p(x) = \frac{341}{256} \cdot \left(\frac{1}{4}\right)^x$, kde $x = 0, 1, 2, 3, 4$ je konkrétna hodnota náhodnej premennej X . V zadaní b) náhodná premenná X nadobúda hodnotu ľubovoľného celého nezáporného čísla x s pravdepodobnosťou $p(x) = \frac{3}{4} \cdot \left(\frac{1}{4}\right)^x$.

Príklad 3.3 Nech náhodná premenná X znamená minimálnu hodnotu padnutých bodov pri hode dvoma hracími kockami. Nájdite zákon rozdelenia pravdepodobnosti danej náhodnej premennej X .

Riešenie. Priestor elementárnych javov znázorníme graficky:

						
	1,1	1,2	1,3	1,4	1,5	1,6
	2,1	2,2	2,3	2,4	2,5	2,6
	3,1	3,2	3,3	3,4	3,5	3,6
	4,1	4,2	4,3	4,4	4,5	4,6
	5,1	5,2	5,3	5,4	5,5	5,6
	6,1	6,2	6,3	6,4	6,5	6,6

Obor hodnôt náhodnej premennej X je množina $M = \{1, 2, 3, 4, 5, 6\}$. Pravdepodobnosti určíme podľa klasickej definície pravdepodobnosti $P(X = 1) = \frac{11}{36}$, $P(X = 2) = \frac{9}{36}$, atď. Výsledky zapíšeme do tabuľky a dostaneme zákon rozdelenia pravdepodobnosti.

Skontrolujeme, či je súčet všetkých pravdepodobností rovný číslu 1.

x_i	1	2	3	4	5	6	Σ
p_i	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{36}{36} = 1$

3.2 DISTRIBUČNÁ FUNKCIA NÁHODNEJ PREMENNÉJ

Nech je X ľubovoľná náhodná premenná (diskrétna alebo spojitá).

Definícia 3.2 Distribučnou funkciou (niekedy aj zákonom rozdelenia pravdepodobnosti) nazývame funkciu $F: R \rightarrow R: \forall x \in R: F(x) = P(X \leq x)$, kde $P(X \leq x)$ je kumulatívna pravdepodobnosť.

Poznámka 3.1 V niektorej literatúra distribučná funkcia sa definuje vzťahom $F(x) = P(X < x)$.

Príklad 3.4 Nech X je diskretná náhodná premenná s oborom hodnôt $M = \{x_1, x_2, x_3, x_4\}$. Vyjadrite distribučnú funkciu náhodnej premennej X , ktorej zákon (rad) rozdelenia pravdepodobnosti je daný tabuľkou:

x_i	x_1	x_2	x_3	x_4	Σ
p_i	p_1	p_2	p_3	p_4	1

Pre kumulatívnu pravdepodobnosť $P(X \leq x)$, t.j. distribučnú funkciu F platí:

$$x < x_1 \quad F(x) = P(X \leq x) = P(\emptyset) = 0$$

$$x_1 \leq x < x_2 \quad F(x) = P(X \leq x) = P(X = x_1) = p_1$$

$$x_2 \leq x < x_3 \quad F(x) = P(X \leq x) = P((X = x_1) \vee (X = x_2)) = P(X = x_1) + P(X = x_2) = p_1 + p_2$$

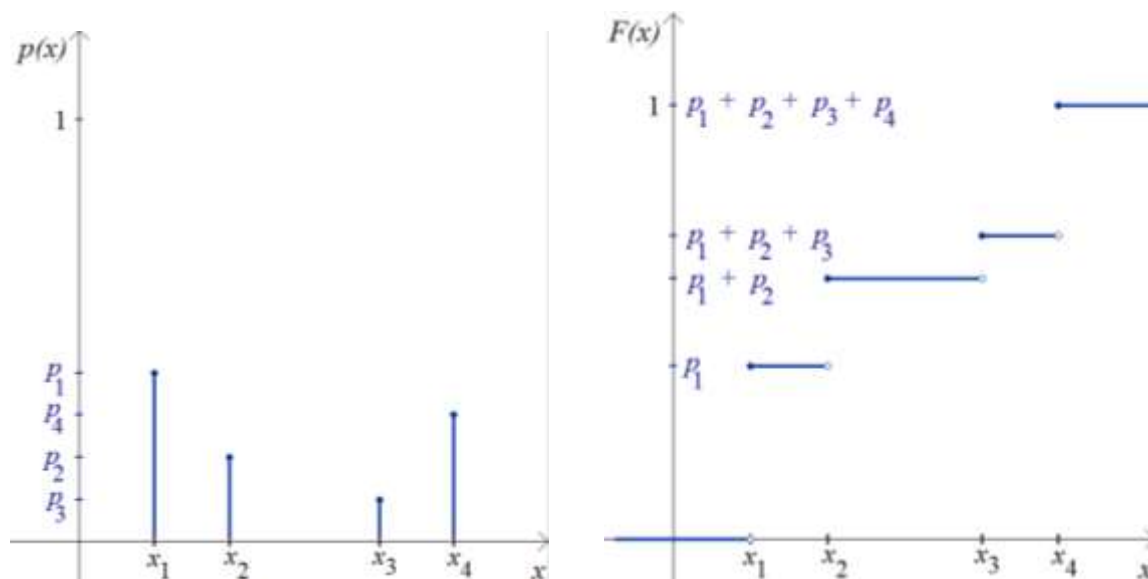
$$x_3 \leq x < x_4 \quad F(x) = P(X \leq x) = P((X = x_1) \vee (X = x_2) \vee (X = x_3)) = P(X = x_1) + P(X = x_2) + P(X = x_3) = p_1 + p_2 + p_3$$

$$x_4 \leq x \quad F(x) = P(X \leq x) = P((X = x_1) \vee (X = x_2) \vee (X = x_3) \vee (X = x_4)) = P(X = x_1) + P(X = x_2) + P(X = x_3) + P(X = x_4) = p_1 + p_2 + p_3 + p_4 = P(I) = 1$$

Z uvedeného vyplýva, že distribučná funkcia diskretnej náhodnej premennej X je daná viacpredpisovou funkciou:

$$F(x) = \begin{cases} 0 & \text{pre } x < x_1; \\ p_1 & \text{pre } x_1 \leq x < x_2; \\ p_1 + p_2 & \text{pre } x_2 \leq x < x_3; \\ p_1 + p_2 + p_3 & \text{pre } x_3 \leq x < x_4; \\ 1 & \text{pre } x_4 \leq x \end{cases}$$

Paličkový graf radu rozdelenia pravdepodobnosti $p(x)$ a graf distribučnej funkcie $F(x)$ danej diskretnej náhodnej premennej X je znázornený na obr. 3.1.



Obr. 3.1 Grafické vyjadrenie vzťahu medzi $p(x)$ a $F(x)$

Ak je náhodná premenná X diskretná, potom jej distribučná funkcia F je vždy „schodovitá“ (neklesajúca) funkcia, ktorá je nespojitá vo všetkých bodoch x_i , ktoré odpovedajú možným hodnotám tejto premennej. Skok v bode x_i sa rovná hodnote $p_i = P(x_i)$. Distribučná funkcia F je sprava spojitá.

Veta 3.1 Vlastnosti distribučnej funkcie:

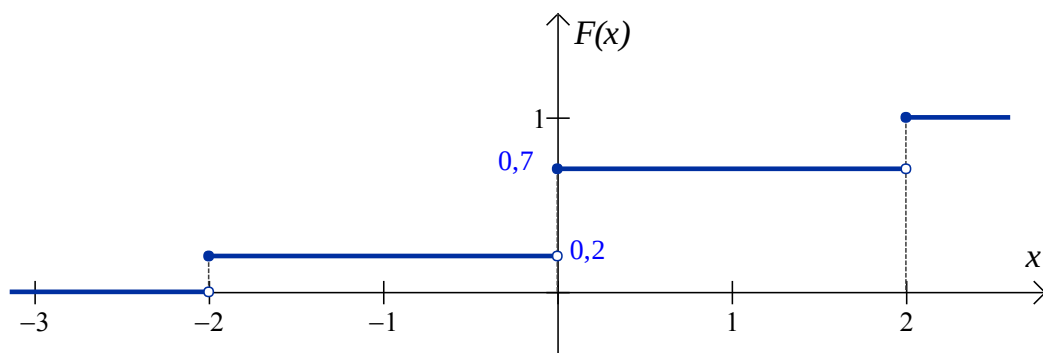
1. pre $\forall x \in \mathbb{R}: 0 \leq F(x) \leq 1$;
2. $\lim_{x \rightarrow -\infty} F(x) = 0$;
3. $\lim_{x \rightarrow \infty} F(x) = 1$;
4. Pre $\forall x_1, x_2 \in \mathbb{R}: x_1 < x_2 \Rightarrow P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$.

Príklad 3.5 Nech X je diskretná náhodná premenná s distribučnou funkciou

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < -2; \\ 0,2 & \text{pre } -2 \leq x < 0; \\ 0,7 & \text{pre } 0 \leq x < 2; \\ 1 & \text{pre } x \geq 2. \end{cases}$$

Vyjadrite zákon rozdelenia pravdepodobnosti takto definovanej náhodnej premennej X pomocou jej pravdepodobnostnej funkcie.

Riešenie. Grafické vyjadrenie zadanej distribučnej funkcie je na obr. 3.2.

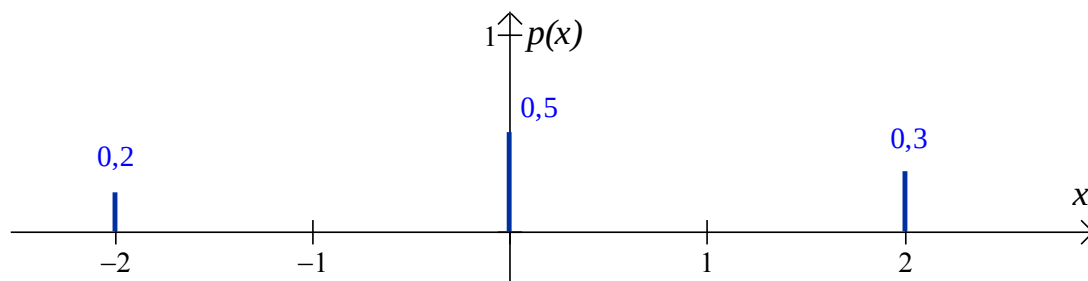


Obr. 3.2 Graf distribučnej funkcie $F(x)$

Skoky v bodoch x_i v distribučnej funkcii $F(x)$ sú rovné hodnotám hľadanej pravdepodobnosti $p(x_i) = P(X = x_i)$, prislúchajúcej konkrétnym hodnotám x_i náhodnej premennej X . Preto vieme zo zadania distribučnej funkcie $F(x)$ vyjadriť pravdepodobnostnú tabuľku s pravdepodobnosťami $p(x)$, prislúchajúcimi jednotlivým hodnotám, t.j. -2, 0, 2, ktoré môže náhodná premenná X nadobudnúť:

x_i	-2	0	2	Σ
p_i	0,2	0,5	0,3	1
$\sum_i p_i$	0,2	0,7	1	-

Grafické znázornenie rozdelenia pravdepodobnosti $p(x)$ je na nasledujúcom obr. 3.3.



Obr. 3.3 Paličkový graf pravdepodobnostnej funkcie $p(x)$

Príklad 3.6 Určte kumulatívne pravdepodobnosti náhodnej premennej X , ktorá znamená minimálnu hodnotu padnutých bodov pri hode dvoma hracími kockami.

Riešenie. Kumulatívne pravdepodobnosti vyjadríme doplnením ďalšieho riadku pravdepodobnostnej tabuľky z príkladu č 3.3.

x_i	1	2	3	4	5	6	Σ
p_i	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{36}{36} = 1$
kumulatívna pravdepodobnosť	$\frac{11}{36}$	$\frac{20}{36}$	$\frac{27}{36}$	$\frac{32}{36}$	$\frac{35}{36}$	$\frac{36}{36}$	

Príklad 3.7 Medzi 40 výrobkami je 5 nepodarkov. Nech náhodná premenná X znamená počet nepodarkov medzi 3 náhodne vybranými výrobkami. Nájdite zákon rozdelenia pravdepodobnosti danej náhodnej premennej X a jej distribučnú funkciu. Aká je pravdepodobnosť toho, že takto definovaná náhodná premenná nadobudne hodnotu menšiu ako 2?

Riešenie. Pri riešení zadanej úlohy v Exceli je možné využiť funkciu =HYPGEOM.DIST. Jej použitie je zobrazené na obr. 3.4.

	A	B	C	D	E	F
1	x_i	0	1	2	3	Σ
2	p_i	=HYPGEOM.DIST(B1;3;5;40;0)	=HYPGEOM.DIST(C1;3;5;40;0)	=HYPGEOM.DIST(D1;3;5;40;0)	=HYPGEOM.DIST(E1;3;5;40;0)	=SUM(B2:E2)
3	kumulatívne pravdepodobnosti	=HYPGEOM.DIST(B1;3;5;40;0)	=B3+C2	=C3+D2	=D3+E2	

	A	B	C	D	E	F
1	x_i	0	1	2	3	Σ
2	p_i	0,66245	0,30111	0,03543	0,00101	1
3	kumulatívne pravdepodobnosti	0,66245	0,96356	0,99899	1	

Obr. 3.4 Výpočet hodnôt $p(x)$ a $F(x)$ realizovaný v Exceli

Pravdepodobnosť toho, že náhodná premenná X nadobudne hodnotu menšiu ako 2 je rovná distribučnej funkcii (resp. kumulatívnej pravdepodobnosti) v čísle 1, preto jej veľkosť bude na základe vytvorenej tabuľky rovná 96,4%.

$$P(X < 2) = P(X \leq 1) = F(1) = 0,96356 \text{ (96,4\%)}$$

Príklad 3.8 Pravdepodobnosť vyrobenia nepodarku je 0,1. Vyrobíme 4 výrobky. Nech náhodná premenná X znamená počet nepodarkov medzi vyrobenými výrobkami. Nájdite zákon (tabuľku) rozdelenia pravdepodobnosti náhodnej premennej X . Koľko nepodarkov vyrobíme s najväčšou pravdepodobnosťou?

Riešenie. Tabuľka – zákon rozdelenia náhodnej premennej X je na obr. 3.5. Na výpočet jednotlivých pravdepodobností v Exceli použijeme funkciu =BINOM.DIST.

	A	B	C	D	E	F	G
1	x_i	0	1	2	3	4	Σ
2	p_i	=BINOM.DIST(B1;4;0,1;0)	=BINOM.DIST(C1;4;0,1;0)	=BINOM.DIST(D1;4;0,1;0)	=BINOM.DIST(E1;4;0,1;0)	=BINOM.DIST(F1;4;0,1;0)	=SUM(B2:F2)

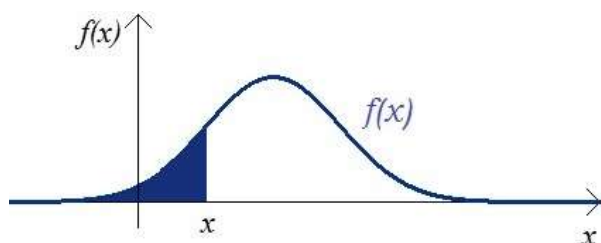
	A	B	C	D	E	F	G
1	x_i	0	1	2	3	4	Σ
2	p_i	0,6561	0,2916	0,0486	0,0036	0,0001	1

Obr. 3.5 Výpočet hodnôt $p(x)$ realizovaný v Exceli

Z vypočítanej tabuľky možno prečítať, že náhodná premenná nadobudne s najväčšou pravdepodobnosťou (65,61%) hodnotu 0, t.j. s najväčšou pravdepodobnosťou nebude medzi vyrobenými 4 výrobkami ani jeden chybný.

3.3 HUSTOTA (ZÁKON ROZDELENIA) SPOJITEJ NÁHODNEJ PREMENEJ

Definícia 3.3 Náhodná premenná X s distribučnou funkciou F sa nazýva spojitá, ak existuje taká integrovateľná funkcia $f(x)$, že pre $\forall x \in \mathbb{R}$ platí: $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$. Funkciu $f(x)$ nazývame hustota rozdelenia pravdepodobnosti spojitej náhodnej premennej X .



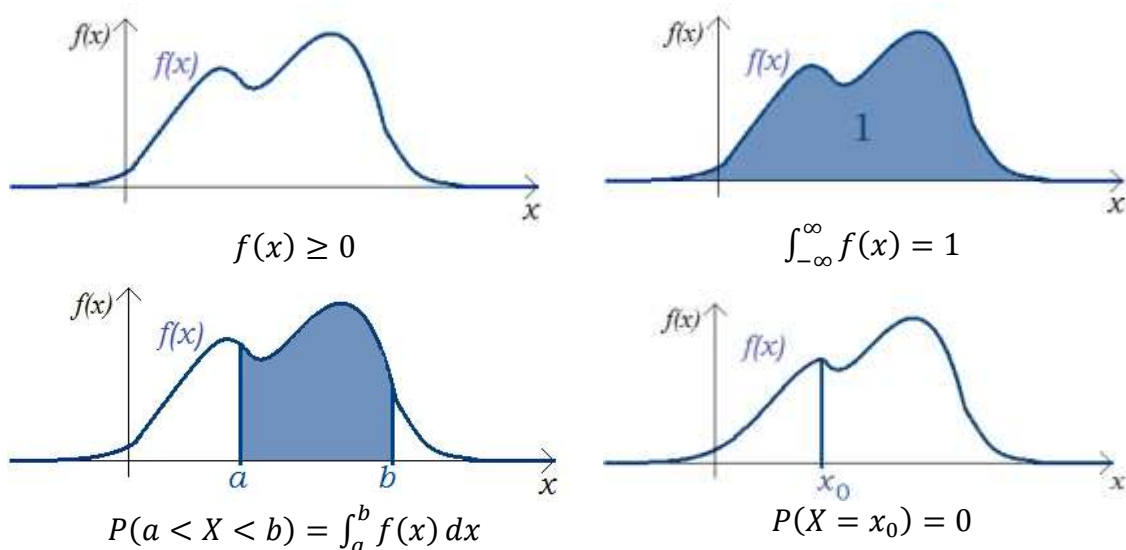
Obr. 3.6 Grafická interpretácia vzťahu $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$

Veta 3.2 Nech X je spojitá náhodná premenná s distribučnou funkciou F . Nech a, b sú reálne čísla. Potom hustota f rozdelenia pravdepodobnosti X má tieto vlastnosti:

1. $f(x) \geq 0$;
2. $\int_{-\infty}^{\infty} f(x) = 1$;
3. $f(x) = F'(x)$ (okrem bodov nespojitosti funkcie f);
4. $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_a^b f(x) dx$.

Poznámka 3.2 Prvá a druhá vlastnosť hustoty je nutnou a postačujúcou podmienkou k tomu, aby funkcia $f(x)$ bola hustotou rozdelenia pravdepodobnosti nejakej spojitej náhodnej premennej X .

Poznámka 3.3 $P(X = x_0) = 0$ ak $x_0 \in \mathbb{R}$.



Obr. 3.7 Grafická interpretácia vlastností hustoty $f(x)$

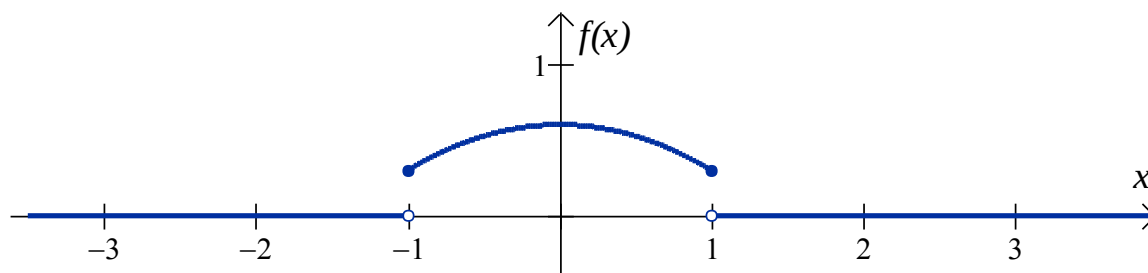
Príklad 3.9 Určte c tak, aby dvojpredpisová funkcia $f(x)$ bola hustotou rozdelenia pravdepodobnosti spojitej náhodnej premennej X : $f(x) = \begin{cases} 0,6 + cx^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}$.

Vypočítajte jej distribučnú funkciu a smerodajnú odchýlku.

Riešenie. Od funkcie $f(x)$ ako hustoty rozdelenia pravdepodobnosti spojitej náhodnej premennej X vyžadujeme, aby bola nezáporná, t.j. jej graf ležal nad x -ovou osou a súčasne, aby plošný obsah pod grafom bol rovný jednej, preto môžeme písať

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{-1} 0 dx + \int_{-1}^1 (0,6 + cx^2) dx + \int_1^{\infty} 0 dx = \left[0,6x + c \cdot \frac{x^3}{3} \right]_{-1}^1 = 0,6 + \frac{c}{3} - \left(-0,6 + c \cdot \frac{(-1)^3}{3} \right) = 1,2 + \frac{2}{3}c \Rightarrow 1 = 1,2 + \frac{2}{3}c \Leftrightarrow c = -0,3.$$

Po dosadení $f(x) = \begin{cases} 0,6 - 0,3x^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}$. Graf takto definovanej funkcie je zobrazený na nasledujúcom obrázku 3.8.



Obr. 3.8 Graf funkcie hustoty rozdelenia pravdepodobnosti náhodnej premennej X

$$f(x) = \begin{cases} 0,6 - 0,3x^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}$$

Distribučná funkcia je definovaná nevlastným integrálom, vyjadrujúcim plošný obsah pod krivkou funkcie hustoty ohraňenej sprava číslom x : $F(x) = \int_{-\infty}^x f(t) dt$.

$$x \in (-\infty, -1): F(x) = \int_{-\infty}^x 0 dt = 0$$

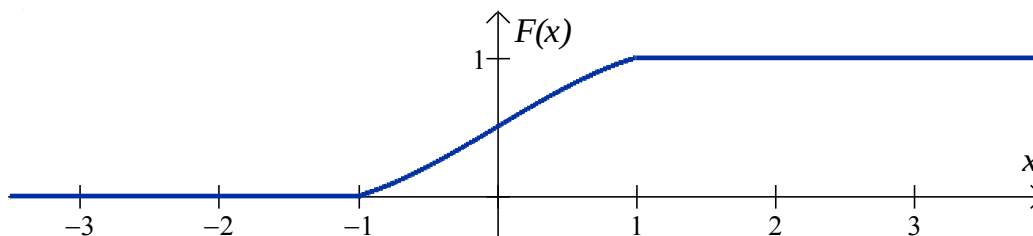
$$x \in \langle -1, 1 \rangle: F(x) = \int_{-\infty}^{-1} 0 dt + \int_{-1}^x (0,6 - 0,3t^2) dt = 0 + \left[0,6t - 0,3 \cdot \frac{t^3}{3} \right]_{-1}^x = 0,6x - 0,1x^3 - (-0,6 + 0,1) = -0,1x^3 + 0,6x + 0,5$$

$$x \in (1, \infty): F(x) = \int_{-\infty}^{-1} 0 dt + \int_{-1}^1 (0,6 - 0,3t^2) dt + \int_1^x 0 dt = \left[0,6t - 0,3 \cdot \frac{t^3}{3} \right]_{-1}^1 = 0,6 - 0,1 - (-0,6 + 0,1) = 1$$

Výsledok stručne zapíšeme:

$$F(x) = \begin{cases} 0 & \text{pre } x \leq -1 \\ -0,1x^3 + 0,6x + 0,5 & \text{pre } -1 \leq x \leq 1 \\ 1 & \text{pre } x \geq 1 \end{cases}$$

Graf distribučnej funkcie je na nasledujúcom obrázku 3.9.



Obr. 3.9 Graf distribučnej funkcie $F(x)$ prislúchajúcej náhodnej premennej X s hustotou

$$\text{pravdepodobnosti } f(x) = \begin{cases} 0,6 - 0,3x^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}$$

3.4 ČÍSELNÉ CHARAKTERISTIKY

Číselné charakteristiky náhodnej premennej sú číselné údaje, ktoré poskytujú základnú predstavu o náhodnej premennej.

Číselné charakteristiky náhodnej premennej možno rozdeliť na skupiny:

- Charakteristiky polohy (stredná hodnota, modus, medián, kvantily);
- Charakteristiky variability (disperzia, smerodajná odchýlka);
- Všeobecné charakteristiky (momenty);
- Charakteristiky symetrie
- Charakteristiky koncentrácie.

3.4.1 Charakteristiky polohy

3.4.1.1 Stredná hodnota – Mean – EXPECTED VALUE (očakávaná hodnota) – $E(X)$

Stredná hodnota náhodnej premennej – označuje sa $E(X)$ je číslo, ktoré opisuje, okolo ktorej hodnoty na číselnej osi sú sústredené samotné hodnoty náhodnej premennej X .

Definícia 3.4 Pre diskrétnu náhodnú premennú stredná hodnota je číslo $E(X) = \sum x_i p_i$, pre spojitú náhodnú premennú je $E(X) = \int_{-\infty}^{\infty} x f(x) dx$ (za podmienky, že rad, resp. integrál existuje).

Veta 3.3 Nech X, Y sú náhodné premenné, $a, b, c \in \mathbb{R}$. Potom vlastnosti strednej hodnoty náhodnej premennej sú:

1. $E(c) = c$;
2. $E(aX + b) = aE(X) + b$;
3. $E(X + Y) = E(X) + E(Y)$.

Príklad 3.10 Nech rozdelenie diskkrétnej náhodnej premennej X je dané tabuľkou:

x_i	-2	0	1	4	6
p_i	0,025	0,1	c	0,3	0,175

Nájdite konštantu c a strednú hodnotu náhodnej premennej X .

Riešenie. Aby daná tabuľka bola zákonom rozdelenia diskkrétnej náhodnej premennej X , musí platiť, že $\sum_{i=1}^5 p_i = 1$. Odtiaľ dostaneme, že $c = 0,4$.

Stredná hodnota diskkrétnej náhodnej premennej X bude číslo $E(X) = \sum_{i=1}^5 x_i \cdot p_i = (-2) \cdot 0,025 + 0 \cdot 0,1 + 1 \cdot 0,4 + 4 \cdot 0,3 + 6 \cdot 0,175 = 2,6$.

Príklad 3.11 Nech je daná hustota rozdelenia spojitej náhodnej premennej X :
 $f(x) = \begin{cases} cx & x \in (0,2); \\ 0 & \text{inak} \end{cases}$. Nájdite konštantu c a strednú hodnotu náhodnej premennej X .

Riešenie. Podľa vlastnosti hustoty rozdelenia spojitej náhodnej premennej nutnou a postačujúcou podmienkou, aby daná funkcia bola hustotou rozdelenia spojitej náhodnej premennej X je, aby $f(x) \geq 0$ a $\int_{-\infty}^{\infty} f(x)dx = 1$.

Z prvej podmienky dostaneme, že $c \geq 0$ a z druhej podmienky máme:

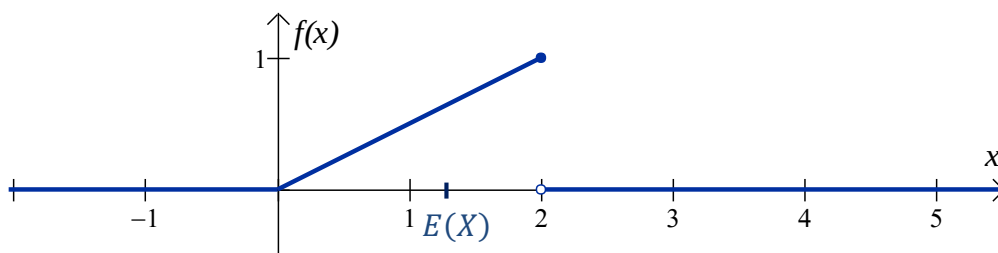
$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^0 f(x)dx + \int_0^2 f(x)dx + \int_2^{\infty} f(x)dx = \int_{-\infty}^0 0dx + \int_0^2 cxdx + \int_2^{\infty} 0dx = \int_0^2 cxdx = \left[\frac{cx^2}{2} \right]_0^2 = 2c = 1, \text{ odkiaľ dostaneme pre konštantu } c = \frac{1}{2}.$$

Teda hustota rozdelenia pravdepodobnosti bude daná funkciou $f(x) = \begin{cases} \frac{1}{2}x & x \in (0,2); \\ 0 & \text{inak} \end{cases}$.

Graf funkcie hustoty rozdelenia pravdepodobnosti náhodnej premennej X je na obr. 3.10.

Stredná hodnota tejto náhodnej premennej X je

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^0 xf(x)dx + \int_0^2 xf(x)dx + \int_2^{\infty} xf(x)dx = \int_0^2 x \frac{1}{2}x dx = \frac{1}{2} \int_0^2 x^2 dx = \frac{1}{2} \left[\frac{x^3}{3} \right]_0^2 = \frac{4}{3}.$$



Obr. 3.10 Graf funkcie hustoty pravdepodobnosti $f(x) = \begin{cases} \frac{1}{2}x & x \in \langle 0, 2 \rangle \\ 0 & \text{inak} \end{cases}$

náhodnej premennej X a jej stredná hodnota $E(X)$

Poznámka 3.3 Pri určovaní strednej hodnoty $E(X)$ hľadáme x -ovú hodnotu ťažiska hmotnej plochy pod krivkou hustoty rozdelenia pravdepodobnosti.

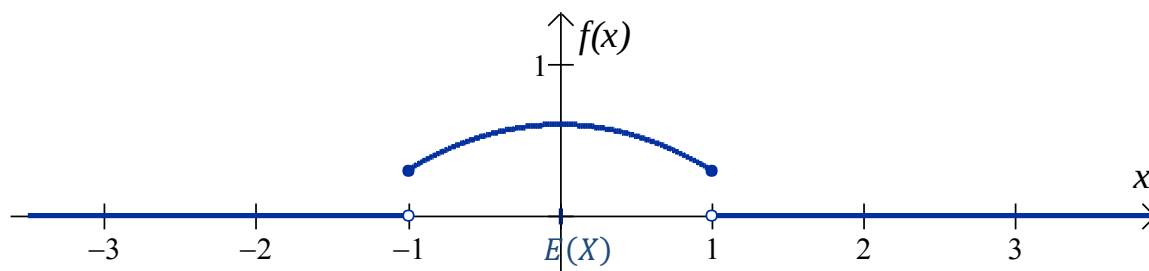
Príklad 3.12 Vypočítajte strednú hodnotu náhodnej premennej X s hustotou pravdepodobnosti

$$f(x) = \begin{cases} 0,6 - 0,3x^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}.$$

Riešenie. Na základe definície strednej hodnoty

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_{-\infty}^{-1} x \cdot 0 dx + \int_{-1}^1 x \cdot (0,6 - 0,3x^2) dx + \int_1^{\infty} x \cdot 0 dx = \int_{-1}^1 (0,6x - 0,3x^3) dx = \\ &= \left[0,6 \frac{x^2}{2} - 0,3 \frac{x^4}{4} \right]_{-1}^1 = 0,3 - \frac{0,3}{4} - \left(0,3 - \frac{0,3 \cdot (-1)^4}{4} \right) = 0. \end{aligned}$$

Strednú hodnotu $E(X)$ vieme určiť aj priamo zo symetrie grafu funkcie hustoty pravdepodobnosti náhodnej premennej X , obr. 3.11.



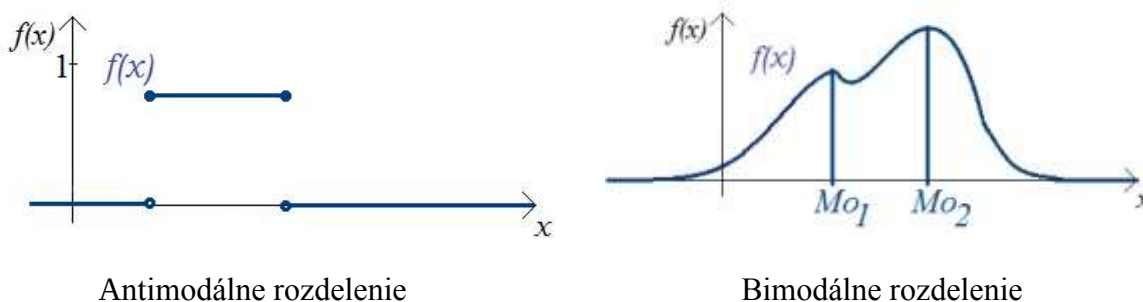
Obr. 3.11 Graf funkcie $f(x) = \begin{cases} 0,6 - 0,3x^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}$ má os symetrie v $x = 0$,

preto stredná hodnota $E(X) = 0$

3.4.1.2 Modus – MODE

Definícia 3.5 Ak X je diskrétna náhodná premenná, potom modus ($Mo(X)$) je tá hodnota náhodnej premennej, ktorá má najväčšiu pravdepodobnosť, t.j. je to najpravdepodobnejšia hodnota X . Ak X je spojitá náhodná premenná, potom jej modus ($Mo(X)$) je číslo, v ktorom má jej funkcia hustoty pravdepodobnosti $f(x)$ ostré lokálne maximum.

Poznámka 3.4 Ak náhodná premenná nemá modus, hovoríme, že náhodná premenná má antimodálne rozdelenie, ak má jeden modus, náhodná premenná má monomodálne rozdelenie, ak má dva modusy, náhodná premenná má bimodálne rozdelenie a ak existujú viac ako dva modusy, hovoríme, že náhodná premenná má polymodálne rozdelenie.



Obr. 3.12 Náhodná premenná nemusí mať vždy iba jeden modus

Príklad 3.13 Nájdite modus náhodných premenných X_1 a X_2 , ak náhodná premenná X_1 je

popísaná pravdepodobnostnou funkciou $p(x) = \begin{cases} 0,6 & \text{pre } x = 1; \\ 0,2 & \text{pre } x = 2; \\ 0,2 & \text{pre } x = 3; \\ 0 & \text{inak;} \end{cases}$ a náhodná premenná X_2

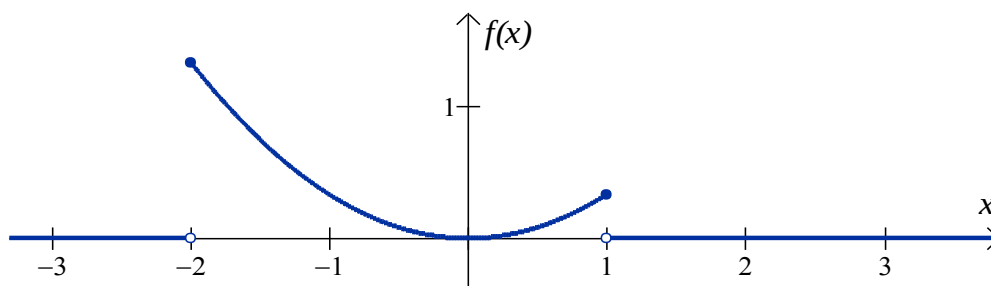
má hustota pravdepodobnosti $f(x) = \begin{cases} \frac{1}{2}x & x \in (0,2); \\ 0 & \text{inak} \end{cases}$.

Riešenie. Náhodná premenná X_1 je diskrétna – jej modus je rovný hodnote 1, ktorá sa vyskytuje s najväčšou pravdepodobnosťou. Modus spojitkej náhodnej premennej X_2 je rovný 2, nakoľko hustota rozdelenia pravdepodobnosti má v čísle 2 ostré lokálne maximum.

Príklad 3.14 Nájdite modus spojitkej náhodnej premennej X s funkciou hustoty

pravdepodobnosti $f(x) = \begin{cases} \frac{1}{3}x^2 & \text{pre } x \in (-2,1); \\ 0 & \text{inak} \end{cases}$.

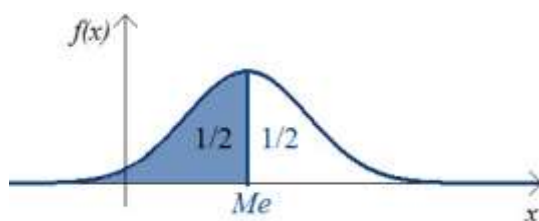
Riešenie. Spojitá náhodná premenná X má bimodálne rozdelenie, jej funkcia hustoty pravdepodobnosti $f(x)$ nadobúda ostré lokálne maximá v bodoch $x = -2$ a $x = 1$, ako možno vidieť na obr. 3.13. Má preto dva modusy $Mo_1(X) = -2$ a $Mo_2(X) = 1$.



Obr. 3.13 Náhodná premenná X má funkciu hustoty $f(x) = \begin{cases} \frac{1}{3}x^2 & x \in \langle -2, 1 \rangle; \\ 0 & \text{inak} \end{cases}$

3.4.1.3 Medián – MEDIAN

Definícia 3.6 Medián $Me(X)$ spojitej náhodnej premennej X je hodnota, pre ktorú platí $F(Me) = P(X \leq Me) = \int_{-\infty}^{Me} f(x)dx = \frac{1}{2}$. Geometrický význam je zobrazený na obr. 3.14.



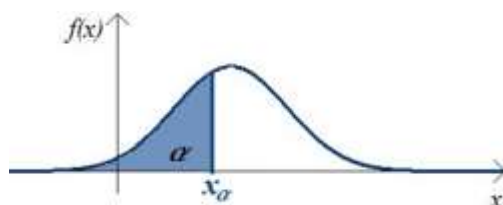
Obr. 3.14 Grafická interpretácia mediánu

3.4.1.4 Kvantily spojitej náhodnej premennej

Definícia 3.7 Nech X je spojitá náhodná premenná, nech $F(x)$ je jej distribučná funkcia, ktorá je rastúca na množine M (podmnožine R). Nech distribučná funkcia zobrazuje množinu M na interval $(0,1)$. Potom inverznú funkciu k tej časti tejto distribučnej funkcie, ktorá je definovaná na množine M , nazývame kvantilovou funkciou spojitej náhodnej premennej X a označujeme $F^{-1}(x)$.

Definícia 3.8 Nech F^{-1} je kvantilová funkcia náhodnej premennej X . Potom číslo $x_\alpha = F^{-1}(\alpha)$, kde $\alpha \in (0,1)$ sa nazýva α -kvantil spojitej náhodnej premennej X .

Poznámka 3.5 Pre α -kvantil náhodnej premennej X platí $x_\alpha = F^{-1}(\alpha) \Leftrightarrow F(x_\alpha) = \alpha$. Preto kvantil $x_{0,5}$ je medián, t.j. $x_{0,5} = Me(X)$.



Obr. 3.15 Grafická interpretácia α -kvantilu

α je pravdepodobnosť „naľavo“ od x_α (veľkosť „zľava ohraničenej“ plochy pod hustotou rozdelenia).

Definícia 3.9 Kvantil $x_{0,25}$ sa nazýva prvý alebo dolný kvartil, $x_{0,5}$ stredný kvartil alebo medián, $x_{0,75}$ tretí alebo horný kvartil náhodnej premennej. Kvantily $x_{0,1}, x_{0,2}, \dots, x_{0,9}$ sa po rade nazývajú prvý decil, druhý decil, ..., deviaty decil. Kvantily $x_{0,01}, x_{0,02}, \dots, x_{0,99}$ sa po rade nazývajú prvý percentil, druhý percentil, ..., deväťdesiaty deviaty percentil.

Príklad 3.15 Nájdite dolný kvartil náhodnej premennej, ktorej hustota je

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

Riešenie. Podľa definície dolný kvartil je $x_{0,25} = F^{-1}(0,25)$. Distribučná funkcie takto

$$\text{definovanej náhodnej premennej je } F(x) = \int_{-\infty}^x \frac{1}{2}t dt = \begin{cases} 0 & x \in (-\infty, 0); \\ \frac{x^2}{4} & x \in \langle 0,2 \rangle; \\ 1 & x \in \langle 2, \infty \rangle. \end{cases}$$

Na intervale $\langle 0,2 \rangle$ existuje k funkcii $F(x)$ inverzná funkcia $F^{-1}(x) = \sqrt{4x}$, $x \in \langle 0,1 \rangle$.

Na základe uvedeného, dolný kvartil $x_{0,25} = F^{-1}(0,25) = \sqrt{4 \cdot 0,25} = 1$.

Poznámka 3.6 Výpočet kvantilov vybraných rozdelení v Exceli je bližšie uvedený na stranách 99, 108, 110 a 113.

3.4.2 Charakteristiky variability

3.4.2.1 Disperzia – rozptyl – VARIANCE – $D(X)$

Charakteristikou variability náhodnej premennej je disperzia, ktorá určuje rozptýlenosť hodnôt náhodnej premennej okolo jej strednej hodnoty. Je definovaná ako priemer druhých mocnín odchýlok od priemeru, t.j. je daná nasledovným vzťahom $D(X) = E[(X - E(X))^2]$.

Pre diskretnú náhodnú premennú X rozptyl sa dá vyjadriť ako $D(X) = \sum_i (x_i - E(X))^2 p_i$

a pre spojitú náhodnú premennú X rozptyl sa rovná $D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$.

Veta 3.4 Nech X je náhodná premenná, kde a, b, c sú konštanty, potom disperzia (rozptyl) má nasledujúce vlastnosti:

1. $D(c) = 0$,
2. $D(aX + b) = a^2 D(X)$,
3. $D(X) = E(X^2) - [E(X)]^2$ (veta o výpočte).

Čím väčšia je disperzia, tým viac kolíšu hodnoty náhodnej premennej okolo strednej hodnoty.

3.4.2.2 Štandardná (smerodajná) odchýlka – STANDARD DEVIATION – $\sigma(X)$

Definícia 3.10 Druhá odmocnina z disperzie náhodnej premennej X sa nazýva smerodajná odchýlka (štandardná odchýlka, stredná kvadratická odchýlka). Označuje sa $\sigma(X)$ a platí $\sigma(X) = \sqrt{D(X)}$.

Príklad 3.16 Diskrétna náhodná premenná X je daná tabuľkou:

x_i	-2	1	3	4	6
p_i	0,1	0,3	0,4	0,15	0,05

Vypočítajte rozptyl a smerodajnú odchýlku náhodnej premennej X .

Riešenie. Pre zjednodušenie výpočtov príklad riešime pomocou Excelu. Vytvoríme funkciu čiastočného súčinu, ktorú kopírujeme do potrebných buniek. Na určenie súčtu môžeme použiť preddefinovanú funkciu =SUM. Pri výpočte disperzie $D(X)$ pomocou definície je pri kopírovaní potrebné uzamknúť bunku, ktorá obsahuje vypočítanú strednú hodnotu $E(X)$ náhodnej premennej X , v našom prípade to bola bunka G3. Smerodajnú odchýlku definovanú ako druhú odmocninu z hodnoty disperzie vypočítame funkciou =SQRT alebo umocnením na $\frac{1}{2}$. Výpočet v Exceli je zobrazený na nasledujúcom obr. 3.16.

	A	B	C	D	E	F	G
1	xi	-2	1	3	4	6	Σ
2	pi	0,1	0,3	0,4	0,15	0,05	=SUM(B2:F2)
3	xi*pi	=B1*B2	=C1*C2	=D1*D2	=E1*E2	=F1*F2	=SUM(B3:F3)
4	xi^2*pi	=B1^2*B2	=C1^2*C2	=D1^2*D2	=E1^2*E2	=F1^2*F2	=SUM(B4:F4)
5	(xi-E(X))^2*pi	=B1-\$G\$3^2*B2	=C1-\$G\$3^2*C2	=D1-\$G\$3^2*D2	=E1-\$G\$3^2*E2	=F1-\$G\$3^2*F2	=SUM(B5:F5)
6							
7	E(X)=	=G3					
8	E(X^2)=	=G4					
9	D(X)=	=G5					
10	D(X)=	=B8-B7^2					
11	σ(X)=	=SQRT(B9)					

- výpočet podľa definície
- výpočet podľa vety o výpočte

	A	B	C	D	E	F	G
1	xi	-2	1	3	4	6	Σ
2	pi	0,1	0,3	0,4	0,15	0,05	1,00
3	xi*pi	-0,2	0,3	1,2	0,6	0,3	2,20
4	xi^2*pi	0,4	0,3	3,6	2,4	1,8	8,50
5	(xi-E(X))^2*pi	1,76	0,43	0,26	0,49	0,72	3,66
6							
7	E(X)=	2,20					
8	E(X^2)=	8,50					
9	D(X)=	3,66					
10	D(X)=	3,66					
11	σ(X)=	1,91					

Obr. 3.16 Výpočet realizovaný v Exceli

Disperzia náhodnej premennej X je $D(X) = 3,66$ a smerodajná odchýlka $\sigma(X) = \sqrt{3,66} = 1,91$.

Príklad 3.17 Nech je daná spojitá náhodná premenná X pomocou funkcie hustoty pravdepodobnosti $f(x) = \begin{cases} -\frac{2}{9}x & \text{pre } x \in \langle -3, 0 \rangle; \\ 0 & \text{inak} \end{cases}$. Nájdite rozptyl a smerodajnú odchýlku náhodnej premennej X .

Riešenie. Najprv musíme vypočítať strednú hodnotu

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{-3} x \cdot 0dx + \int_{-3}^0 x \cdot \left(-\frac{2}{9}\right)xdx + \int_0^{\infty} x \cdot 0dx = -\frac{2}{9} \int_0^2 x^2 dx = -\frac{2}{9} \left[\frac{x^3}{3}\right]_{-3}^0 = -\frac{2}{27} (0^3 - (-3)^3) = -\frac{2}{27} \cdot 27 = -2.$$

Rozptyl (disperzia) $D(X)$ spojitej náhodnej premennej X je definovaný ako $D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$, preto

$$D(X) = \int_{-\infty}^{\infty} (x - (-2))^2 f(x)dx = \int_{-\infty}^{-3} (x+2)^2 \cdot 0dx + \int_{-3}^0 (x+2)^2 \cdot \left(-\frac{2}{9}\right)xdx + \int_0^{\infty} (x+2)^2 \cdot 0dx = -\frac{2}{9} \int_{-3}^0 (x+2)^2 \cdot xdx = -\frac{2}{9} \int_{-3}^0 (x^2 + 4x + 4) \cdot xdx = -\frac{2}{9} \int_{-3}^0 (x^3 + 4x^2 + 4x)dx = -\frac{2}{9} \left[\frac{x^4}{4} + 4\frac{x^3}{3} + 4\frac{x^2}{2}\right]_{-3}^0 = -\frac{2}{9} \left(0 - \left(\frac{(-3)^4}{4} + 4 \cdot \frac{(-3)^3}{3} + 2 \cdot (-3)^2\right)\right) = \frac{1}{2}.$$

Disperziu môžeme vypočítať aj pomocou tretej vlastnosti disperzie, t.j. vety o výpočte $D(X) = E(X^2) - [E(X)]^2$.

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{-\infty}^{-3} x^2 \cdot 0dx + \int_{-3}^0 x^2 \cdot \left(-\frac{2}{9}\right)xdx + \int_0^{\infty} x^2 \cdot 0dx = -\frac{2}{9} \int_{-3}^0 x^3 dx = -\frac{2}{9} \left[\frac{x^4}{4}\right]_{-3}^0 = -\frac{1}{18} (0^4 - (-3)^4) = \frac{9}{2}$$

$$D(X) = \frac{9}{2} - (-2)^2 = \frac{1}{2}.$$

Príklad 3.18 Nájdite rozptyl a smerodajnú odchýlku náhodnej premennej X z príkladu 3. 12.

Riešenie. Funkcie hustoty pravdepodobnosti spojitej náhodnej premennej X je

$$f(x) = \begin{cases} 0,6 - 0,3x^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}. \text{ Jej stredná hodnota } E(X) = 0. \text{ Podľa definície rozptyl}$$

$$D(X) = \int_{-\infty}^{\infty} (x - E(x))^2 \cdot f(x)dx = \int_{-\infty}^{-1} (x-0)^2 \cdot 0dx + \int_{-1}^1 (x-0)^2 \cdot (0,6 - 0,3x^2)dx + \int_1^{\infty} (x-0)^2 \cdot 0dx = \int_{-1}^1 (0,6x^2 - 0,3x^4)dx = \left[0,6 \cdot \frac{x^3}{3} - 0,3 \cdot \frac{x^5}{5}\right]_{-1}^1 = (0,2 - 0,06 - (-0,2 - (-0,06))) = 0,28.$$

Potom smerodajná odchýlka $\sigma(X) = \sqrt{D(X)} = \sqrt{0,28} \doteq 0,53$.

3.4.3 Všeobecné charakteristiky – momenty

Definícia 3.11 Nech k je prirodzené číslo. Nech X je diskretná náhodná premenná, potom k -ty začiatkový moment premennej X sa definuje ako $z_k(X) = \sum_i x_i^k \cdot p_i$, ak X je spojitá náhodná premenná s hustotou f , potom k -ty začiatkový moment premennej X sa definuje ako $z_k(X) = \int_{-\infty}^{\infty} x^k f(x)dx$.

Definícia 3.12 Nech k je prirodzené číslo. Nech X je diskrétna náhodná premenná, potom k -ty centrálny moment premennej X sa definuje ako $c_k(X) = \sum_i (x_i - E(X))^k \cdot p_i$, ak X je spojitá náhodná premenná s hustotou f , potom k -ty centrálny moment premennej X sa definuje ako $c_k(X) = \int_{-\infty}^{\infty} (x - E(X))^k f(x) dx$.

Veta 3.5 Nech existujú k -te začiatočné a centrálné momenty náhodnej premennej X , $k \in N$. Potom centrálné momenty sa dajú vyjadriť pomocou začiatočných:

1. $c_1 = 0$;
2. $c_2 = z_2 - z_1^2$;
3. $c_3 = z_3 - 3z_2z_1 + 2z_1^3$;
4. $c_4 = z_4 - 4z_3z_1 + 6z_2z_1^2 - 3z_1^4$.

Príklad 3.19 Diskrétna náhodná premenná X je daná tabuľkou:

x_i	-2	1	3	4	6
p_i	0,1	0,3	0,4	0,15	0,05

Vypočítajte prvé štyri začiatočné a centrálné momenty náhodnej premennej X .

Riešenie. Pri riešení v Exceli je efektívne vytvoriť vlastné funkcie a použiť ich kopírovanie do vybraných buniek – obr. 3.17. Pri kopírovaní využijeme uzamknutie stĺpcov B a C, v ktorých sú zadané hodnoty x_i a k nim prislúchajúce pravdepodobnosti p_i .

	A	B	C	D	E	F	G	H	I	J	K
1		x_i	p_i	$x_i \cdot p_i$	$x_i^2 \cdot p_i$	$x_i^3 \cdot p_i$	$x_i^4 \cdot p_i$	$(x_i - E(X)) \cdot p_i$	$(x_i - E(X))^2 \cdot p_i$	$(x_i - E(X))^3 \cdot p_i$	$(x_i - E(X))^4 \cdot p_i$
2		-2	0,1	=B2*C2	=B2^2*C2	=B2^3*C2	=B2^4*C2	=(B2-\$D\$10)*C2	=(B2-\$D\$10)^2*C2	=(B2-\$D\$10)^3*C2	=(B2-\$D\$10)^4*C2
3		1	0,3	=B3*C3	=B3^2*C3	=B3^3*C3	=B3^4*C3	=(B3-\$D\$10)*C3	=(B3-\$D\$10)^2*C3	=(B3-\$D\$10)^3*C3	=(B3-\$D\$10)^4*C3
4		3	0,4	=B4*C4	=B4^2*C4	=B4^3*C4	=B4^4*C4	=(B4-\$D\$10)*C4	=(B4-\$D\$10)^2*C4	=(B4-\$D\$10)^3*C4	=(B4-\$D\$10)^4*C4
5		4	0,15	=B5*C5	=B5^2*C5	=B5^3*C5	=B5^4*C5	=(B5-\$D\$10)*C5	=(B5-\$D\$10)^2*C5	=(B5-\$D\$10)^3*C5	=(B5-\$D\$10)^4*C5
6		6	0,05	=B6*C6	=B6^2*C6	=B6^3*C6	=B6^4*C6	=(B6-\$D\$10)*C6	=(B6-\$D\$10)^2*C6	=(B6-\$D\$10)^3*C6	=(B6-\$D\$10)^4*C6
7		Σ	=SUM(C3:C6)	=SUM(D3:D6)	=SUM(E3:E6)	=SUM(F3:F6)	=SUM(G3:G6)	=SUM(H3:H6)	=SUM(I3:I6)	=SUM(J3:J6)	=SUM(K3:K6)
8											
9				z def.				z def.	podľa vety		
10		$z_1(X) = E(X)$	=SUM(D3:D6)	2,2				$c_1(X) =$	=SUM(H3:H6)	0	
11		$z_2(X) =$	=SUM(E3:E6)					$c_2(X) = D(X) =$	=SUM(I3:I6)	=D11-D10^2	
12		$z_3(X) =$	=SUM(F3:F6)					$c_3(X) =$	=SUM(J3:J6)	=D12-3*D11*D10+2*D10^3	
13		$z_4(X) =$	=SUM(G3:G6)					$c_4(X) =$	=SUM(K3:K6)	=D13-4*D12*D10+6*D11*D10^2-3*D10^4	

	A	B	C	D	E	F	G	H	I	J	K
1		x_i	p_i	$x_i \cdot p_i$	$x_i^2 \cdot p_i$	$x_i^3 \cdot p_i$	$x_i^4 \cdot p_i$	$(x_i - E(X)) \cdot p_i$	$(x_i - E(X))^2 \cdot p_i$	$(x_i - E(X))^3 \cdot p_i$	$(x_i - E(X))^4 \cdot p_i$
2		-2	0,1	-0,2	0,4	-0,8	1,6	-0,42	1,764	-7,4088	31,1696
3		1	0,3	0,3	0,3	0,3	0,3	-0,36	0,432	-0,5184	0,62208
4		3	0,4	1,2	3,6	10,8	32,4	0,32	0,256	0,2048	0,16384
5		4	0,15	0,6	2,4	9,6	38,4	0,27	0,486	0,8748	1,57464
6		6	0,05	0,3	1,8	10,8	64,8	0,19	0,722	2,7436	10,42568
7		Σ	1	2,2	8,5	30,7	137,5	0	3,66	-4,104	43,9032
8											
9				z def.				z def.	podľa vety		
10		$z_1(X) = E(X)$	2,2	2,2				$c_1(X) =$	0	0	
11		$z_2(X) =$	8,5					$c_2(X) = D(X) =$	3,66	3,66	
12		$z_3(X) =$	30,7					$c_3(X) =$	-4,104	-4,104	
13		$z_4(X) =$	137,5					$c_4(X) =$	43,9032	43,9032	

Obr. 3.17 Výpočet realizovaný v Exceli

Príklad 3.20 Nech je daná spojitá náhodná premenná X pomocou funkcie hustoty rozdelenia pravdepodobnosti $f(x) = \begin{cases} \frac{3}{16}x^2 & \text{pre } x \in \langle -2, 2 \rangle \\ 0 & \text{pre } x \notin \langle -2, 2 \rangle \end{cases}$. Vypočítajte prvé štyri začiatočné a centrálné momenty náhodnej premennej X .

Riešenie. Najprv vypočítame začiatočné momenty.

$$z_1(X) = \int_{-\infty}^{\infty} x^1 \cdot \frac{3}{16} x^2 dx = \frac{3}{16} \int_{-2}^2 x^3 dx = \frac{3}{16} \left[\frac{x^4}{4} \right]_{-2}^2 = 0$$

$$z_2(X) = \int_{-\infty}^{\infty} x^2 \cdot \frac{3}{16} x dx = \frac{3}{16} \int_{-2}^2 x^4 dx = \frac{3}{16} \left[\frac{x^5}{5} \right]_{-2}^2 = \frac{12}{15}$$

$$z_3(X) = \int_{-\infty}^{\infty} x^3 \cdot \frac{3}{16} x dx = \frac{3}{16} \int_{-2}^2 x^5 dx = \frac{3}{16} \left[\frac{x^6}{6} \right]_{-2}^2 = 0$$

$$z_4(X) = \int_{-\infty}^{\infty} x^4 \cdot \frac{3}{16} x dx = \frac{3}{16} \int_{-2}^2 x^5 dx = \frac{3}{16} \left[\frac{x^6}{6} \right]_{-2}^2 = \frac{48}{49}.$$

Centrálné momenty vypočítame podľa vety 3.5 pomocou začiatočných momentov.

$$c_1(X) = 0$$

$$c_2(X) = z_2(X) - z_1^2(X) = \frac{12}{15}$$

$$c_3(X) = z_3(X) - 3z_2(X)z_1(X) + 2z_1^3(X) = 0 - 3 \cdot \frac{12}{15} \cdot 0 + 2 \cdot 0^3 = 0$$

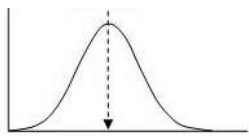
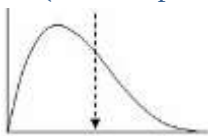
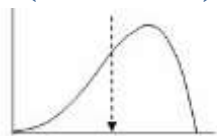
$$c_4(X) = z_4(X) - 4z_3(X)z_1(X) + 6z_2(X)z_1^2(X) - 3z_1^4(X) = \frac{48}{49} - 4 \cdot 0 \cdot 0 + 6 \cdot \frac{12}{15} \cdot 0^2 - 3 \cdot 0^4 = \frac{48}{49}.$$

3.4.4 Charakteristiky symetrie

Koeficient šikmosti – skewness

Koeficient šikmosti sa označuje $\alpha(X)$ a charakterizuje symetriu rozdelenia náhodnej premennej X . Definuje sa ako podiel tretieho centrálneho momentu a tretej mocniny smerodajnej odchýlky náhodnej premennej. Je to teda číslo $\alpha(X) = \frac{c_3}{\sigma^3(X)} = \frac{c_3}{\sqrt{D^3(X)}}$.

Geometrická interpretácia koeficientu šikmosti a jeho vzťahu so strednou hodnotou, modusom a mediánom pre spojitú náhodnú premennú je v nasledujúcej tabuľke:

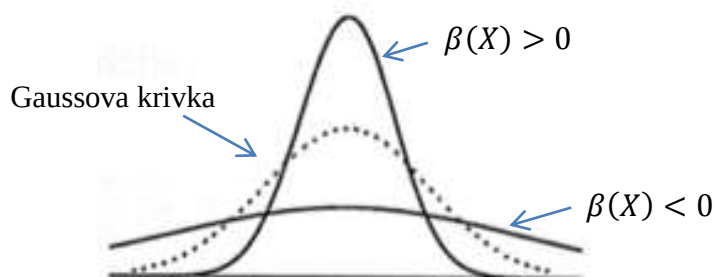
Graf funkcie hustoty je			
	symetrický	nesymetrický	
		zošikmenie doprava (vrchol vpredu) 	zošikmenie doľava (vrchol vzadu) 
koeficient šikmosti $\alpha(X)$ je	0	> 0 (kladný)	< 0 (záporný)
vzťah medzi $E(X)$, Mo , Me	$E(X) = Mo = Me$	$Mo < Me < E(X)$ (častejší výskyt menších hodnôt)	$E(X) < Me < Mo$ (častejší výskyt väčších hodnôt)

3.4.5 Charakteristiky koncentrácie

Koeficient špicatosti – kurtosis

Koeficient špicatosti charakterizuje špicatosť, strmosť rozdelenia náhodnej premennej X , označuje sa $\beta(X)$ a je definovaný vzťahom $\beta(X) = \frac{c_4}{\sigma^4(X)} - 3 = \frac{c_4}{D^2(X)} - 3$.

Koeficient špicatosti porovnáva špicatosť grafu hustoty rozdelenia pravdepodobnosti, resp. polygónu so špicatosťou krivky normálneho (Gaussovoho) rozdelenia s tou istou strednou hodnotou a disperziou, akú má porovnávaná náhodná premenná.



Obr. 3.18 Geometrická interpretácia koeficientu špicatosti

Špicatost' slúži ako miera koncentrácie hodnôt náhodnej premennej okolo jej strednej hodnoty.

Príklad 3.21 Vypočítajte koeficient šikmosti a špicatosti diskkrétnej náhodnej premennej X danej pravdepodobnostnou tabuľkou:

x_i	1	2	3		4	5
p_i	0,05	0,1	0,25		0,4	0,2

Riešenie. Riešenie v Exceli je zobrazené na nasledujúcom obr. 3.19.

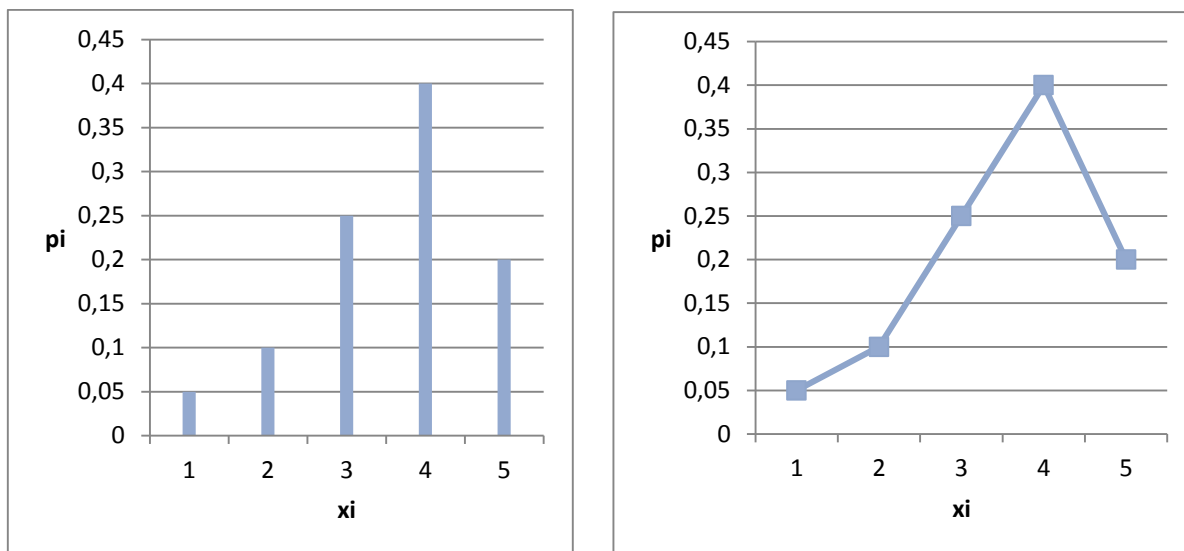
	A	B	C	D	E	F
1	x_i	p_i	$x_i \cdot p_i$	$(x_i - E(X))^2 \cdot p_i$	$(x_i - E(X))^3 \cdot p_i$	$(x_i - E(X))^4 \cdot p_i$
2	1	0,05	=A2*B2	=(A2-\$B\$8)^2*\$B2	=(A2-\$B\$8)^3*\$B2	=(A2-\$B\$8)^4*\$B2
3	2	0,1	=A3*B3	=(A3-\$B\$8)^2*\$B3	=(A3-\$B\$8)^3*\$B3	=(A3-\$B\$8)^4*\$B3
4	3	0,25	=A4*B4	=(A4-\$B\$8)^2*\$B4	=(A4-\$B\$8)^3*\$B4	=(A4-\$B\$8)^4*\$B4
5	4	0,4	=A5*B5	=(A5-\$B\$8)^2*\$B5	=(A5-\$B\$8)^3*\$B5	=(A5-\$B\$8)^4*\$B5
6	5	0,2	=A6*B6	=(A6-\$B\$8)^2*\$B6	=(A6-\$B\$8)^3*\$B6	=(A6-\$B\$8)^4*\$B6
7		=SUM(B2:B6)	=SUM(C2:C6)	=SUM(D2:D6)	=SUM(E2:E6)	=SUM(F2:F6)
8	$E(X)=$	=SUM(C2:C6)				
9	$D(X)=$	=SUM(D2:D6)				
10	$\sigma(X)=$	=SQRT(B9)				
11	$\alpha(X)=$	=E7/B10^3				
12	$\beta(X)=$	=F7/B10^4-3				

↓

	A	B	C	D	E	F
1	x_i	p_i	$x_i \cdot p_i$	$(x_i - E(X))^2 \cdot p_i$	$(x_i - E(X))^3 \cdot p_i$	$(x_i - E(X))^4 \cdot p_i$
2	1	0,05	0,05	0,338	-0,8788	2,28488
3	2	0,1	0,2	0,256	-0,4096	0,65536
4	3	0,25	0,75	0,09	-0,054	0,0324
5	4	0,4	1,6	0,064	0,0256	0,01024
6	5	0,2	1	0,392	0,5488	0,76832
7		1	3,6	1,14	-0,768	3,7512
8	$E(X)=$	3,6				
9	$D(X)=$	1,14				
10	$\sigma(X)=$	1,06771				
11	$\alpha(X)=$	-0,63096				
12	$\beta(X)=$	-0,11357				

Obr. 3.19 Výpočet realizovaný v Exceli

Tvar rozdelenia pravdepodobnosti a jeho súvislosť s hodnotou koeficientu šikmosti a špicatosti môžeme vidieť na nasledujúcom obr. 3.20. Koeficient šikmosti $\alpha(X) = -0,63096$ nám hovorí, že zošikmenie je negatívne, t.j. je ľavostranne zošikmené. Koeficient špicatosti $\beta(X) = -0,11357$ je záporný, a preto možno tvrdiť, že rozdelenie pravdepodobnosti skúmanej diskkrétnej náhodnej premennej X je plochejšie ako Gaussova krivka.



Obr. 3.20 Grafické zobrazenie rozdelenia pravdepodobnosti náhodnej premennej X pomocou „paličkového“ grafu a polygónu – v Exceli zobrazené pomocou ponuky Vložiť grafy

Príklad 3.22 Diskrétna náhodná premenná X je daná pravdepodobnostnou tabuľkou:

x_i	-3	0	1	3	4
p_i	0,1	0,4	0,3	0,15	0,05

Vypočítajte jej charakteristiky variability, symetrie a koncentrácie.

Riešenie.

Disperzia $D(X) = 2,93$, smerodajná odchýlka $\sigma(X) = 1,71$, koeficient šikmosti $\alpha(X) = -0,23$, koeficient špicatosti $\beta(X) = 0,35$.

Riešenie v Exceli je zobrazené na nasledujúcom obr. 3.21.

	A	B	C	D	E
1	x_i	p_i	$(x_i - E(X))^2 \cdot p_i$	$(x_i - E(X))^3 \cdot p_i$	$(x_i - E(X))^4 \cdot p_i$
2	-3	0,1	$=(\$A2-\$B\$8)^2*\$B2$	$=(\$A2-\$B\$8)^3*\$B2$	$=(\$A2-\$B\$8)^4*\$B2$
3	0	0,4	$=(\$A3-\$B\$8)^2*\$B3$	$=(\$A3-\$B\$8)^3*\$B3$	$=(\$A3-\$B\$8)^4*\$B3$
4	1	0,3	$=(\$A4-\$B\$8)^2*\$B4$	$=(\$A4-\$B\$8)^3*\$B4$	$=(\$A4-\$B\$8)^4*\$B4$
5	3	0,15	$=(\$A5-\$B\$8)^2*\$B5$	$=(\$A5-\$B\$8)^3*\$B5$	$=(\$A5-\$B\$8)^4*\$B5$
6	4	0,05	$=(\$A6-\$B\$8)^2*\$B6$	$=(\$A6-\$B\$8)^3*\$B6$	$=(\$A6-\$B\$8)^4*\$B6$
7	$=SUM(B2:B6)$		$=SUM(C2:C6)$	$=SUM(D2:D6)$	$=SUM(E2:E6)$
8	$E(X) = SUMPRODUCT(A2:A6;B2:B6)$				
9	$D(X) = SUM(C2:C6)$				
10	$\sigma(X) = SQRT(B9)$				
11	$\alpha(X) = D7/B10^3$				
12	$\beta(X) = E7/B10^4-3$				

	A	B	C	D	E
1	x_i	p_i	$(x_i - E(X))^2 \cdot p_i$	$(x_i - E(X))^3 \cdot p_i$	$(x_i - E(X))^4 \cdot p_i$
2	-3	0,1	1,33225	-4,8627125	17,74890063
3	0	0,4	0,169	-0,10985	0,0714025
4	1	0,3	0,03675	0,0128625	0,004501875
5	3	0,15	0,828375	1,94668125	4,574700938
6	4	0,05	0,561125	1,87976875	6,297225313
7		1	2,9275	-1,13325	28,69673125
8	$E(X) = 0,65$				
9	$D(X) = 2,9275$				
10	$\sigma(X) = 1,71099$				
11	$\alpha(X) = -0,22625$				
12	$\beta(X) = 0,34841$				

Obr. 3.21 Výpočet realizovaný v Exceli

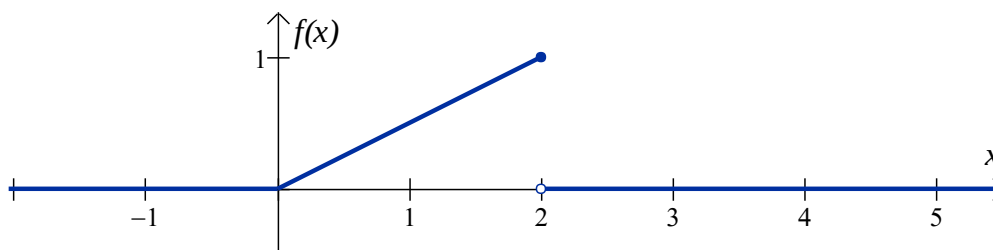
Príklad 3.23 Nech $f(x) = \begin{cases} cx & \text{pre } x \in \langle 0, 2 \rangle; \\ 0 & \text{pre } x \notin \langle 0, 2 \rangle \end{cases}$ je hustota rozdelenia pravdepodobnosti spojitej náhodnej premennej X . Nájdite číslo c , modus, disperziu a smerodajnú odchýlku náhodnej premennej X . Nájdite a znázornite distribučnú funkciu náhodnej premennej X . Nájdite a graficky znázornite nasledujúce kvantily: medián, $x_{0,04}$, $x_{0,77}$.

Riešenie.

1. Aby sme funkciu $f(x)$ mohli nazvať hustotou rozdelenia pravdepodobnosti náhodnej premennej X , musí pre ňu platiť:

- $\int_{-\infty}^{\infty} f(x) dx = 1$;
- pre $\forall x \in R: f(x) \geq 0$.

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^0 0 dx + \int_0^2 cx dx + \int_2^{\infty} 0 dx = \int_0^2 cx dx = c \left[\frac{x^2}{2} \right]_0^2 = 2c \Rightarrow c = \frac{1}{2}$$



Obr. 3.22 Graf funkcie $f(x) = \begin{cases} \frac{1}{2}x & \text{pre } x \in \langle 0, 2 \rangle; \\ 0 & \text{pre } x \notin \langle 0, 2 \rangle \end{cases}$, ktorá je hustotou rozdelenia pravdepodobnosti spojitej náhodnej premennej X .

Vidíme, že je splnená aj druhá podmienka, t.j. že $f(x) \geq 0$ pre $\forall x \in R$ (žiadna časť grafu funkcie $f(x)$ neleží pod x -ovou osou).

2. Modus spojitej náhodnej premennej X $Mo(X)$ je definovaný ako číslo, v ktorom funkcia hustoty rozdelenia pravdepodobnosti tejto spojitej náhodnej premennej X nadobúda svoje ostré lokálne maximum. Z grafu funkcie $f(x)$ na obrázku 3. 22 vidíme, že funkcia má hľadaný extrém v čísle $x = 2$, preto $Mo(X) = 2$.

3. Disperziou alebo rozptylom $D(X)$ spojitej náhodnej premennej X nazývame druhý centrálny moment $c_2(X)$, t.j. počítame $c_k(X) = \int_{-\infty}^{\infty} (x - E(X))^k f(x) dx$ pre $k = 2$. Pri výpočte k -teho centrálného momentu však potrebujeme vždy poznať strednú hodnotu $E(X)$ náhodnej premennej X . Tá je definovaná ako prvý začiatkový moment $p_1(X)$. Preto ako prvé počítame k -ty začiatkový moment $z_k(X) = \int_{-\infty}^{\infty} x^k f(x) dx$ pre $k = 1$:

$$\begin{aligned} E(X) = z_1(X) &= \int_{-\infty}^{\infty} x^1 f(x) dx = \int_{-\infty}^0 x \cdot 0 dx + \int_0^2 x \cdot \frac{1}{2} x dx + \int_2^{\infty} x \cdot 0 dx = \int_0^2 \frac{1}{2} x^2 dx \\ &= \frac{1}{2} \left[\frac{x^3}{3} \right]_0^2 = \frac{1}{6} (2^3 - 0) = \frac{4}{3}. \end{aligned}$$

Potom disperzia:

$$\begin{aligned} D(X) = c_2(X) &= \int_{-\infty}^{\infty} \left(x - \frac{4}{3} \right)^2 f(x) dx \\ &= \int_{-\infty}^0 \left(x - \frac{4}{3} \right)^2 \cdot 0 dx + \int_0^2 \left(x - \frac{4}{3} \right)^2 \cdot \frac{1}{2} x dx + \int_2^{\infty} \left(x - \frac{4}{3} \right)^2 \cdot 0 dx \\ &= \int_0^2 \left(x - \frac{4}{3} \right)^2 \cdot \frac{1}{2} x dx = \frac{1}{2} \int_0^2 \left(x^2 - \frac{8}{3} x + \frac{16}{9} \right) \cdot x dx = \frac{1}{2} \int_0^2 \left(x^3 - \frac{8}{3} x^2 + \frac{16}{9} x \right) dx \\ &= \frac{1}{2} \left[\frac{x^4}{4} - \frac{8}{3} \frac{x^3}{3} + \frac{16}{9} \frac{x^2}{2} \right]_0^2 = \frac{1}{2} \left(4 - \frac{64}{9} + \frac{32}{9} \right) = \frac{1}{2} \cdot \frac{36 - 64 + 32}{9} = \frac{2}{9}. \end{aligned}$$

Poznámka 3.7 Na výpočet disperzie je možné tiež použiť vzťah $D(X) = E(X^2) - [E(X)]^2$.

$E(X^2)$ počítame ako druhý začiatkový moment:

$$\begin{aligned} E(X^2) = z_2(X) &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-\infty}^0 x^2 \cdot 0 dx + \int_0^2 x^2 \cdot \frac{1}{2} x dx + \int_2^{\infty} x^2 \cdot 0 dx = \int_0^2 \frac{1}{2} x^3 dx \\ &= \frac{1}{2} \left[\frac{x^4}{4} \right]_0^2 = \frac{1}{2} \left(\frac{16}{4} - 0 \right) = 2 \end{aligned}$$

$$D(X) = 2 - \left[\frac{4}{3} \right]^2 = \frac{2}{9}.$$

$$4. \text{ Smerodajná odchýlka } s = \sqrt{D(X)} = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3} \doteq 0,4714.$$

5. Distribučná funkcia $F(x)$ náhodnej premennej X je definovaná pre všetky reálne čísla x ako nevlastný integrál $F(x) = \int_{-\infty}^x f(t) dt$.

$$x \in (-\infty, 0): F(x) = \int_{-\infty}^x 0 dt = 0$$

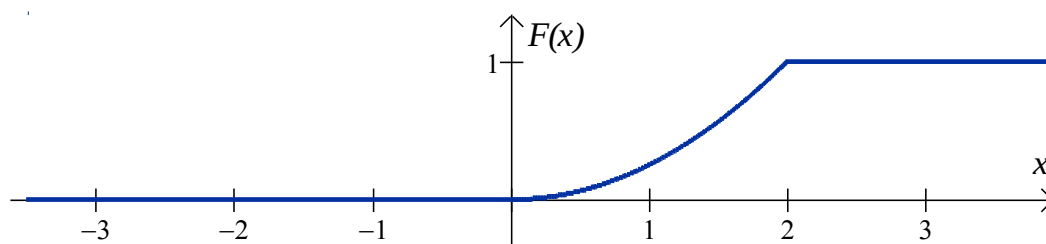
$$x \in \langle 0, 2 \rangle: F(x) = \int_{-\infty}^0 0 dt + \int_0^x \frac{1}{2}t dt = 0 + \frac{1}{2} \left[\frac{t^2}{2} \right]_0^x = \frac{1}{4}(x^2 - 0) = \frac{x^2}{4}$$

$$x \in (2, \infty): F(x) = \int_{-\infty}^0 0 dt + \int_0^2 \frac{1}{2}t dt + \int_2^x 0 dt = 0 + \frac{1}{2} \left[\frac{t^2}{2} \right]_0^2 = \frac{1}{4}(2^2 - 0) = 1$$

Vypočítané zapíšeme nasledovným spôsobom:

$$F(x) = \begin{cases} 0 & \text{pre } x \in (-\infty, 0); \\ \frac{x^2}{4} & \text{pre } x \in \langle 0, 2 \rangle; \\ 1 & \text{pre } x \in (2, \infty). \end{cases}$$

Možno konštatovať, táto trojpredpisová funkcia je na celom definičnom obore, t.j. na množine R spojitá, nezáporná, neklesajúca a jej obor hodnôt $F(x) \in \langle 0, 1 \rangle$ pre $\forall x \in R$. Jej graf môžeme vidieť na nasledujúcom obrázku 3.23.



Obr. 3.23 Graf distribučnej funkcie $F(x) = \begin{cases} 0 & \text{pre } x \in (-\infty, 0); \\ \frac{x^2}{4} & \text{pre } x \in \langle 0, 2 \rangle; \\ 1 & \text{pre } x \in (2, \infty); \end{cases}$ prislúchajúcej hustote rozdelenia pravdepodobnosti spojitý náhodnej premennej X s prepisom $f(x) = \begin{cases} \frac{1}{2}x & \text{pre } x \in \langle 0, 2 \rangle; \\ 0 & \text{pre } x \notin \langle 0, 2 \rangle. \end{cases}$

6. Výpočet kvantilov: $x_{0,5}$; $x_{0,04}$; $x_{0,77}$

Najznámejším kvantilom je medián Me , ktorý označujeme tiež $x_{0,5}$. Jeho hodnota delí interval hodnôt spojitý náhodnej premennej na dva rovnako pravdepodobné intervaly a plochu, ktorá je ohraničená grafom hustoty pravdepodobnosti a x -ovou osou na dve časti s rovnakým plošným obsahom. Platí $P(X \leq x_{0,5}) = P(X \geq x_{0,5}) = \frac{1}{2}$, čo je vyjadrené pomocou funkcie hustoty

pravdepodobnosti ako $\int_{-\infty}^{Me} f(x) dx = \int_{Me}^{\infty} f(x) dx = \frac{1}{2}$, resp. priamo pomocou distribučnej funkcie

$$F(x_{0,5}) = \frac{1}{2}.$$

$$\int_0^{x_{0,5}} \frac{1}{2} x dx = \frac{1}{2}$$

$$\frac{1}{4} [x^2]_0^{x_{0,5}} = \frac{1}{2} \quad / \cdot 4$$

$$x_{0,5}^2 = 2$$

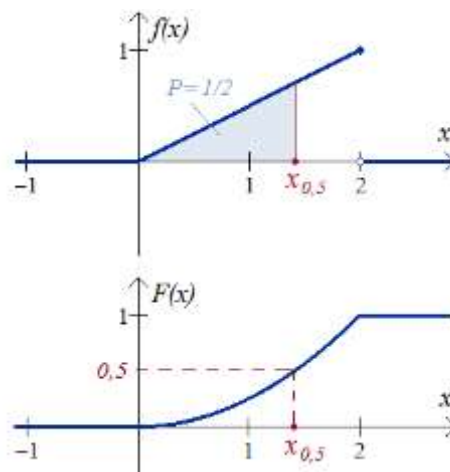
$$x_{0,5} = \pm \sqrt{2}$$

resp.

$$F(x_{0,5}) = \frac{x_{0,5}^2}{4} = \frac{1}{2}$$

$$x_{0,5}^2 = 2 \Rightarrow x_{0,5} = \pm \sqrt{2}$$

Hodnotu $x_{0,5} = -\sqrt{2}$ vylúčime, keďže $-\sqrt{2} \notin \langle 0, 2 \rangle$!



Obr. 3.24 Zobrazenie mediánu $x_{0,5} = \sqrt{2}$ na grafe funkcie hustoty pravdepodobnosti $f(x)$ a na grafe distribučnej funkcie $F(x)$

Aj kvantily $x_{0,04}$ a $x_{0,77}$ (t.j. štvrtý a sedemdesiaty siedmy percentil) je možné vyjadriť pomocou funkcie hustoty pravdepodobnosti, resp. pomocou distribučnej funkcie:

$$\int_0^{x_{0,04}} \frac{1}{2} x dx = \frac{4}{100}$$

$$\frac{1}{4} [x^2]_0^{x_{0,04}} = \frac{4}{100} \quad / \cdot 4$$

$$x_{0,04}^2 = \frac{16}{100}$$

$$x_{0,04}^2 = 0,16$$

$$x_{0,04} = \pm \sqrt{0,16} = \pm 0,4$$

resp.

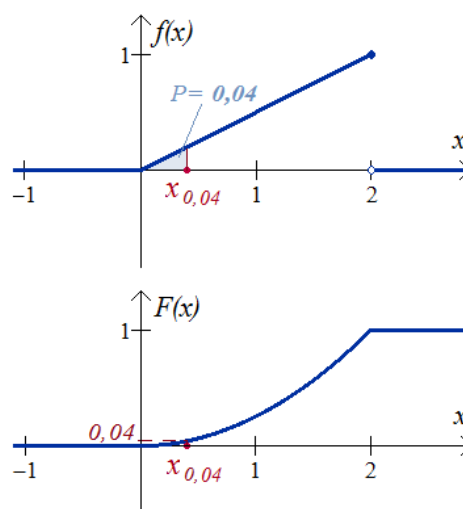
$$F(x_{0,04}) = \frac{x_{0,04}^2}{4} = 0,04$$

$$x_{0,04}^2 = 0,16$$

$$x_{0,04} = \pm \sqrt{0,16} = \pm 0,4$$

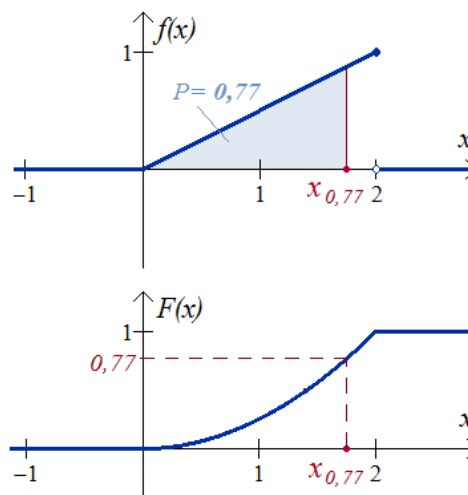
Pozor!

$$-0,4 \notin \langle 0, 2 \rangle \Rightarrow x_{0,04} = 0,4$$



Obr. 3.25 Zobrazenie 4. percentilu $x_{0,04} = 0,4$ na grafe funkcie hustoty pravdepodobnosti $f(x)$ a na grafe distribučnej funkcie $F(x)$

$$\begin{aligned}
\int_0^{x_{0,77}} \frac{1}{2} x dx &= \frac{77}{100} \\
\frac{1}{4} [x^2]_0^{x_{0,77}} &= \frac{77}{100} \quad / \cdot 4 \\
x_{0,77}^2 &= \frac{308}{100} \\
x_{0,77}^2 &= 3,08 \\
x_{0,77} &= \sqrt{3,08} \doteq 1,755 \\
-1,755 &\notin \langle 0, 2 \rangle
\end{aligned}$$



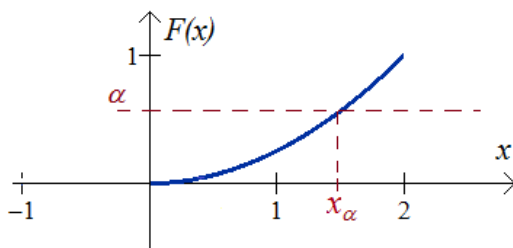
Obr. 3.26 Zobrazenie 77. percentilu $x_{0,77} = 1,755$ na grafe funkcie hustoty pravdepodobnosti $f(x)$ a na grafe distribučnej funkcie $F(x)$

Kvantily je možné určiť tiež pomocou kvantilovej funkcie F^{-1} , ktorá je inverznou funkciou k distribučnej funkcii F , nakoľko hodnoty $F^{-1}(\alpha)$ pre $\alpha \in (0, 1)$ sú definované ako $\alpha \cdot 100\%$ -né kvantily.

Výpočet kvantilovej funkcie je nasledovný:

Keďže inverzná funkcia existuje iba k funkcii, ktorá je prostá, t.j. k funkcii, ktorá má tú vlastnosť, že k jej dvom rôznym hodnotám argumentu prislúchajú vždy dve rôzne funkčné hodnoty, musíme vybrať iba tú časť distribučnej funkcie F , ktorá túto vlastnosť spĺňa. Volíme

$F(x) = \frac{x^2}{4}$ pre $x \in (0, 2)$ – na danom intervale je funkcia rastúca, teda prostá. Z grafu môžeme vidieť, že táto parciálna funkcia zobrazuje interval $(0, 2)$ na interval $(0, 1)$. Nakoľko platí $H(F) = D(F^{-1})$ a $D(F) = H(F^{-1})$, hľadáme kvantilovú funkciu $F^{-1}(\alpha) = ?$ pre $\alpha \in (0, 1)$. Parameter α je známy a volíme ho z intervalu $(0, 1)$, t.j. robíme tzv. α rez na zvolenej časti distribučnej funkcie, obr. 3.27.

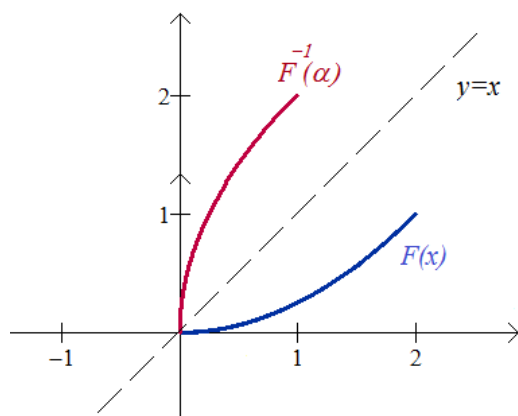


Obr. 3.27 Zobrazenie α -rezu distribučnej funkcie $F(x) = \frac{x^2}{4}$ definovanej pre $x \in (0, 2)$

Kvantilovú funkciu $F^{-1}(\alpha)$ určíme riešením rovnice $\alpha = F(x_\alpha) = \frac{x_\alpha^2}{4}$ s neznámou x_α , pričom x_α je α -kvantil rozdelenia pravdepodobnosti náhodnej premennej X a $\alpha \in (0, 1)$.

$$\alpha = \frac{x_\alpha^2}{4} \Rightarrow x_\alpha = \sqrt{4\alpha}, \text{ nakoľko } -\sqrt{4\alpha} \notin \langle 0, 2 \rangle.$$

Porovnaním definičných oborov a oborov hodnôt vieme rozhodnúť, že $F^{-1}(\alpha) = \sqrt{4\alpha}$ pre $\alpha \in (0, 1)$. Graf kvantilovej funkcie $F^{-1}(\alpha) = \sqrt{4\alpha}$ pre $\alpha \in (0, 1)$ môžeme vidieť na nasledujúcom obr. 3.28.



Obr. 3.28 Zobrazenie grafov distribučnej a kvantilovej funkcie, ktoré sú navzájom inverzné

Pre jednotlivé kvantily potom platí:

$$x_{0,5} = Me = F^{-1}(0,5) = \sqrt{4 \cdot 0,5} = \sqrt{2};$$

$$x_{0,04} = F^{-1}(0,04) = \sqrt{4 \cdot 0,04} = \sqrt{0,16} = 0,4;$$

$$x_{0,77} = F^{-1}(0,77) = \sqrt{4 \cdot 0,77} = \sqrt{3,08} = 1,755.$$

Príklad 3.24 Nech $f(x) = \begin{cases} cx & \text{pre } x \in \langle -3, 0 \rangle; \\ 0 & \text{pre } x \notin \langle -3, 0 \rangle. \end{cases}$ je hustota rozdelenia pravdepodobnosti spojitkej náhodnej premennej X . Nájdite číslo c , modus, strednú hodnotu, disperziu a smerodajnú odchýlku náhodnej premennej X . Nájdite funkčný predpis distribučnej a kvantilovej funkcie náhodnej premennej X a zobrazte ich graficky. Nájdite a graficky znázornite medián. Určte pravdepodobnosť toho, že náhodná premenná X nadobúda hodnoty z intervalu $\langle -2, -1 \rangle$.

Riešenie.

1. Funkcia $f(x)$ musí spĺňať nasledujúce vlastnosti:

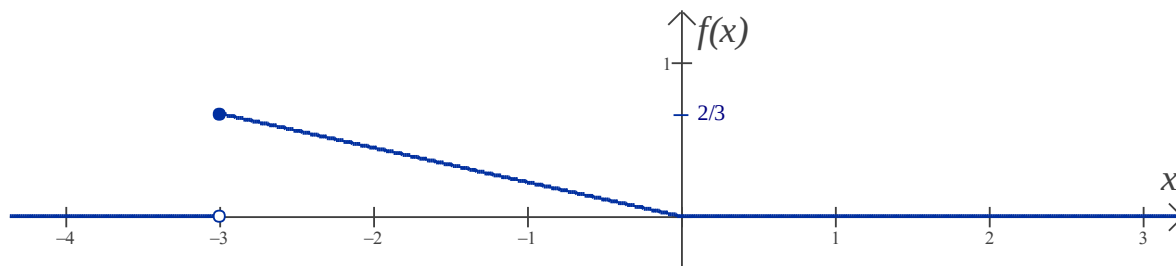
a) $\int_{-\infty}^{\infty} f(x) dx = 1;$

b) pre $\forall x \in R: f(x) \geq 0.$

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{-3} 0 dx + \int_{-3}^0 cx dx + \int_0^{\infty} 0 dx = \int_{-3}^0 cx dx = c \left[\frac{x^2}{2} \right]_{-3}^0 = -\frac{9}{2}c$$

$$\Rightarrow c = -\frac{2}{9}.$$

Zapíšeme $f(x) = \begin{cases} -\frac{2}{9}x & \text{pre } x \in \langle -3, 0 \rangle; \\ 0 & \text{inak} \end{cases}$. Platí aj $f(x) \geq 0$ pre $\forall x \in R$, t.j. žiadna časť grafu funkcie $f(x)$ neleží pod x -ovou osou.



Obr. 3.29 Graf funkcie hustoty rozdelenia pravdepodobnosti spojitej náhodnej premennej X

Poznámka 3.8 Plošný obsah plochy pod priamkou vyjadrujúcou hustotu pravdepodobnosti na intervale $\langle -3, 0 \rangle$, t.j. plošný obsah trojuholníka so základňou $a = 3$ a výškou $v_a = \frac{2}{3}$ je rovný $S = \frac{a \cdot v_a}{2} = \frac{3 \cdot \frac{2}{3}}{2} = 1$.

Pomocou grafu funkcie hustoty pravdepodobnosti vieme nájsť modus $Mo(X)$. Vidíme, že funkcia $f(x)$ nadobúda svoje ostré lokálne maximum v čísle $x = -3$, preto $Mo(X) = -3$.

Stredná hodnota $E(X)$ náhodnej premennej X je definovaná ako prvý začiatkový moment $z_1(X)$:

$$\begin{aligned} E(X) = z_1(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{-3} x \cdot 0 dx + \int_{-3}^0 x \cdot \left(-\frac{2}{9}\right) x dx + \int_0^{\infty} x \cdot 0 dx = -\frac{2}{9} \int_0^2 x^2 dx \\ &= -\frac{2}{9} \left[\frac{x^3}{3} \right]_{-3}^0 = -\frac{2}{27} (0^3 - (-3)^3) = -\frac{2}{27} \cdot 27 = -2. \end{aligned}$$

Poznámka 3.9 Hodnota $E(X)$ musí ležať v intervale, v ktorom je funkcia hustoty pravdepodobnosti nenulová, t.j. $E(X) = -2 \in \langle -3, 0 \rangle$.

Disperzia $D(X)$ spojitej náhodnej premennej X je definovaná ako druhý centrálny moment $c_2(X)$. Počítame $c_2(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$.

$$\begin{aligned} D(X) = c_2(X) &= \int_{-\infty}^{\infty} (x - (-2))^2 f(x) dx = \\ &= \int_{-\infty}^{-3} (x + 2)^2 \cdot 0 dx + \int_{-3}^0 (x + 2)^2 \cdot \left(-\frac{2}{9}\right) x dx + \int_0^{\infty} (x + 2)^2 \cdot 0 dx = \end{aligned}$$

$$\begin{aligned}
&= -\frac{2}{9} \int_{-3}^0 (x+2)^2 \cdot x dx = -\frac{2}{9} \int_{-3}^0 (x^2 + 4x + 4) \cdot x dx = -\frac{2}{9} \int_{-3}^0 (x^3 + 4x^2 + 4x) dx \\
&= -\frac{2}{9} \left[\frac{x^4}{4} + 4 \frac{x^3}{3} + 4 \frac{x^2}{2} \right]_{-3}^0 = -\frac{2}{9} \left(0 - \left(\frac{(-3)^4}{4} + 4 \cdot \frac{(-3)^3}{3} + 2 \cdot (-3)^2 \right) \right) = \frac{1}{2}
\end{aligned}$$

Disperzia vypočítaná pomocou začiatkových momentov: $D(X) = E(X^2) - [E(X)]^2$.

$$\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-\infty}^{-3} x^2 \cdot 0 dx + \int_{-3}^0 x^2 \cdot \left(-\frac{2}{9}\right) x dx + \int_0^{\infty} x^2 \cdot 0 dx = -\frac{2}{9} \int_{-3}^0 x^3 dx = \\
&= -\frac{2}{9} \left[\frac{x^4}{4} \right]_{-3}^0 = -\frac{1}{18} (0^4 - (-3)^4) = \frac{9}{2}
\end{aligned}$$

Potom $D(X) = \frac{9}{2} - (-2)^2 = \frac{1}{2}$.

Poznámka 3.10 Disperzia (rozptyl) náhodnej premennej X musí vyjsť vždy kladné číslo.

Smerodajná (štandardná) odchýlka $s = \sqrt{D(x)} = \sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}} \cdot \frac{\sqrt{2}}{\sqrt{2}} = \frac{\sqrt{2}}{2}$.

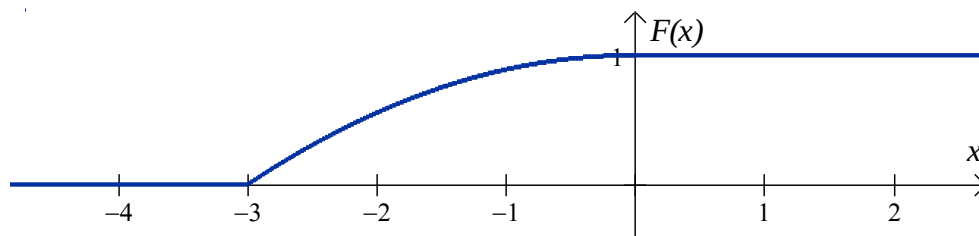
Distribučná funkcia $F(x)$ náhodnej premennej X je definovaná pre všetky reálne čísla x ako nevlastný integrál $F(x) = \int_{-\infty}^x f(t) dt$.

Pre $x \in (-\infty, -3)$: $F(x) = 0$.

Pre $x \in \langle -3, 0 \rangle$: $F(x) = \int_{-\infty}^{-3} 0 dt + \int_{-3}^x \left(-\frac{2}{9}\right) t dt = -\frac{2}{9} \int_{-3}^x t dt = -\frac{2}{9} \left[\frac{t^2}{2} \right]_{-3}^x =$
 $-\frac{1}{9} (x^2 - (-3)^2) = -\frac{1}{9} (x^2 - 9) = 1 - \frac{x^2}{9}$.

Pre $x \in (0, \infty)$: $F(x) = 1$.

Stručne zapíšeme $F(x) = \begin{cases} 0 & \text{pre } x \in (-\infty, -3); \\ 1 - \frac{x^2}{9} & \text{pre } x \in \langle -3, 0 \rangle; \\ 1 & \text{pre } x \in (0, \infty) \end{cases}$ a zobrazíme graficky:



Obr. 3.30 Graf distribučnej funkcie $F(x)$

Poznámka 3.11 Distribučná funkcia je vždy na celom svojom definičnom obore, t.j. na množine $\mathbb{R}$ spojitá, nezáporná a neklesajúca a jej obor hodnôt je $\langle 0, 1 \rangle$.

Kvantilová funkciu F^{-1} je inverznou funkciou k distribučnej funkcii F . Keďže inverzná funkcia existuje iba k funkcii, ktorá je prostá, t.j. k funkcii, ktorá má tú vlastnosť, že k jej dvom rôznym hodnotám argumentu prislúchajú vždy dve rôzne funkčné hodnoty, musíme vybrať iba tú časť distribučnej funkcie F , ktorá túto vlastnosť spĺňa. Volíme $F(x) = 1 - \frac{x^2}{9}$ pre $x \in \langle -3, 0 \rangle$ – na danom intervale je funkcia rastúca, teda prostá. Z grafu $F(x)$ môžeme vidieť, že táto parciálna funkcia zobrazuje interval $\langle -3, 0 \rangle$ na interval $\langle 0, 1 \rangle$. Pretože platí $H(F) = D(F^{-1})$ a $D(F) = H(F^{-1})$, inverzná funkcia k $F(x)$, na intervale $\langle -3, 0 \rangle$, nazývaná kvantilová funkcia $F^{-1}(\alpha)$, bude definovaná pre $\alpha \in \langle 0, 1 \rangle$.

$F: y = 1 - \frac{x^2}{9}$ je prostá na $\langle -3, 0 \rangle \Rightarrow$ existuje k nej inverzná F^{-1}

$$D(F) = \langle -3, 0 \rangle = H(F^{-1})$$

$$H(F) = \langle 0, 1 \rangle = D(F^{-1})$$

$$F^{-1}: x = 1 - \frac{y^2}{9}$$

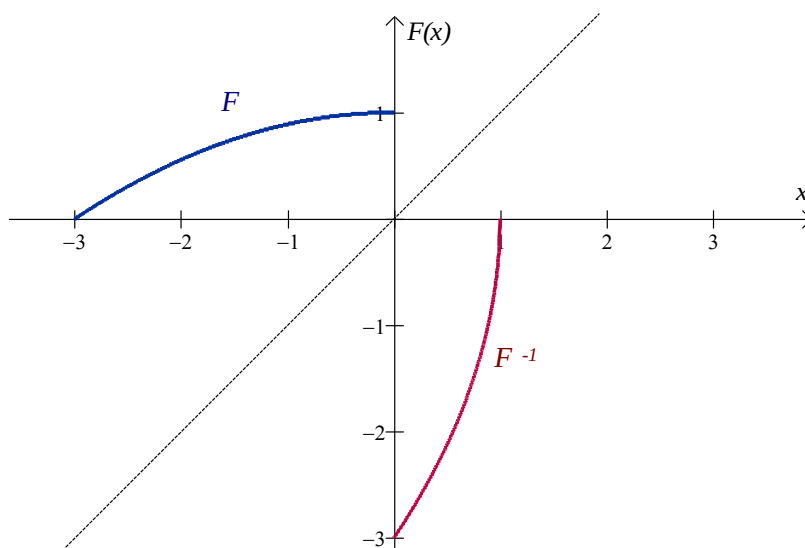
$$9x = 9 - y^2$$

$$y^2 = 9 - 9x$$

odkiaľ

$$y = \pm\sqrt{9 - 9x}$$

Porovnaním definičných oborov a oborov hodnôt vieme rozhodnúť, že kvantilová funkcia je $F^{-1}(\alpha) = -\sqrt{9 - 9\alpha}$ pre $\alpha \in \langle 0, 1 \rangle$.

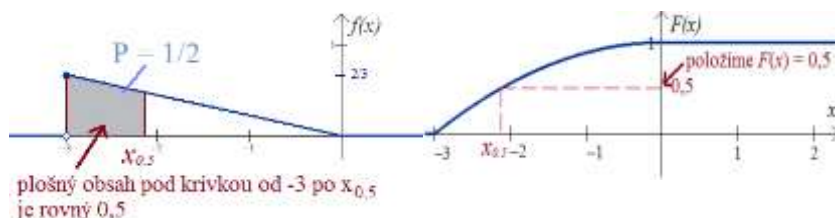


Obr. 3.31 Zobrazenie kvantilovej funkcie $F^{-1}(\alpha) = -\sqrt{9 - 9\alpha}$, $\alpha \in \langle 0, 1 \rangle$

Medián Me značíme tiež $x_{0,5}$ (značí tiež piaty decil aj päťdesiaty percentil).

Nasledujú tri spôsoby výpočtu:

Me vyjadrený pomocou plošného obsahu pod krivkou hustoty pravdepodobnosti $f(x)$:

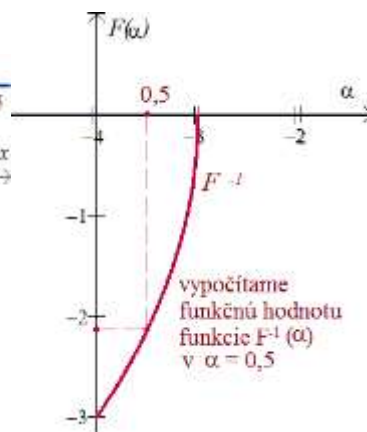


$$\begin{aligned} \int_{-3}^{x_{0,5}} \left(-\frac{2}{9}\right)x dx &= \frac{1}{2} \\ -\frac{2}{9} \left[\frac{x^2}{2} \right]_{-3}^{x_{0,5}} &= \frac{1}{2} \quad / \cdot 2 \\ -\frac{2}{9} (x_{0,5}^2 - (-3)^2) &= 1 \\ x_{0,5}^2 - 9 &= -\frac{9}{2} \\ x_{0,5}^2 &= \frac{9}{2} \\ x_{0,5} &= -\frac{3}{\sqrt{2}} \\ (\text{hodnota } \frac{3}{\sqrt{2}} &\notin \langle -3, 0 \rangle). \end{aligned}$$

Me vyjadrený pomocou distribučnej funkcie $F(x)$:

$$\begin{aligned} \frac{1}{2} &= 1 - \frac{x_{0,5}^2}{9} \\ \frac{x_{0,5}^2}{9} &= 1 - \frac{1}{2} \\ x_{0,5} &= -\frac{3}{\sqrt{2}} \\ (\text{hodnota } \frac{3}{\sqrt{2}} &\notin \langle -3, 0 \rangle). \end{aligned}$$

Me vyjadrený pomocou kvantilovej funkcie $F(x)$:



$$\begin{aligned} F^{-1}\left(\frac{1}{2}\right) &= -\sqrt{9 - 9 \cdot \frac{1}{2}} \\ F^{-1}\left(\frac{1}{2}\right) &= -\frac{3}{\sqrt{2}}. \end{aligned}$$

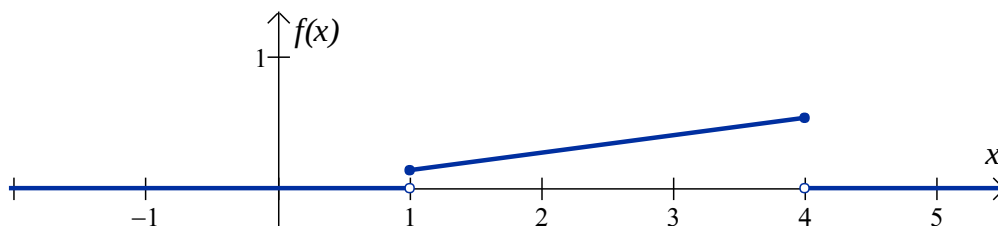
Poznámka 3.12 Medián Me , rovnako ako akýkoľvek kvantil, musí ležať v intervale, v ktorom je funkcia hustoty pravdepodobnosti nenulová, t.j. v intervale $\langle -3, 0 \rangle$.

Pravdepodobnosť toho, že náhodná premenná X nadobúda hodnoty z intervalu $\langle -2, -1 \rangle$, môžeme vypočítať pomocou hustoty pravdepodobnosti alebo pomocou distribučnej funkcie. Platí $P(a \leq X \leq b) = \int_a^b f(x) dx = [F(x)]_a^b = F(b) - F(a)$. Potom $P(-2 \leq X \leq -1) = \int_{-2}^{-1} -\frac{2}{9}x dx = -\frac{2}{9} \left[\frac{x^2}{2} \right]_{-2}^{-1} = -\frac{1}{9} [(-1)^2 - (-2)^2] = \frac{1}{3}$ alebo $P(-2 \leq X \leq -1) = F(-1) - F(-2) = \left[1 - \frac{(-1)^2}{9} \right] - \left[1 - \frac{(-2)^2}{9} \right] = \frac{1}{3}$.

Príklad 3.25 Nech $f(x) = \begin{cases} cx & \text{pre } 1 \leq x \leq 4; \\ 0 & \text{inak} \end{cases}$ je hustota rozdelenia pravdepodobnosti spojitej náhodnej premennej X . Nájdite číslo c . Nájdite funkčný predpis distribučnej funkcie náhodnej premennej X a zobrazte jej graf. Určte prvý decil a zobrazte ho na grafe distribučnej funkcie. Aká je pravdepodobnosť, že náhodná premenná X nadobúda hodnoty od 2 po 3?

Riešenie. $1 = \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^1 0 dx + \int_1^4 cx dx + \int_4^{\infty} 0 dx = \int_1^4 cx dx = c \left[\frac{x^2}{2} \right]_1^4 = \frac{15}{2} c \Rightarrow c = \frac{2}{15}.$

Po dosadení $f(x) = \begin{cases} \frac{2}{15}x & \text{pre } x \in \langle 1, 4 \rangle; \\ 0 & \text{pre } x \notin \langle 1, 4 \rangle \end{cases}.$



Obr. 3.32 Graf funkcie hustoty pravdepodobnosti spojitej náhodnej premennej X

Určíme distribučnú funkciu $F(x)$.

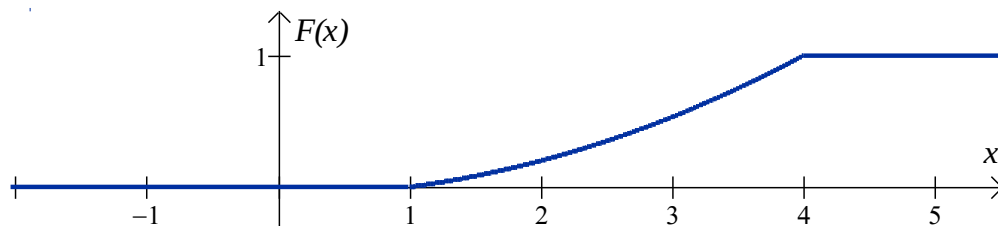
Pre $\forall x \in \mathbb{R}$ platí $F(x) = \int_{-\infty}^x f(t) dt$.

$$x \in (-\infty, 1): F(x) = \int_{-\infty}^x 0 dt = 0$$

$$x \in \langle 1, 4 \rangle: F(x) = \int_{-\infty}^1 0 dt + \int_1^x \frac{2}{15} t dt = 0 + \frac{2}{15} \left[\frac{t^2}{2} \right]_1^x = \frac{1}{15} (x^2 - 1)$$

$$x \in (4, \infty): F(x) = \int_{-\infty}^1 0 dt + \int_1^4 \frac{2}{15} t dt + \int_4^x 0 dt = 0 + \frac{2}{15} \left[\frac{t^2}{2} \right]_1^4 = \frac{1}{15} (4^2 - 1) = 1$$

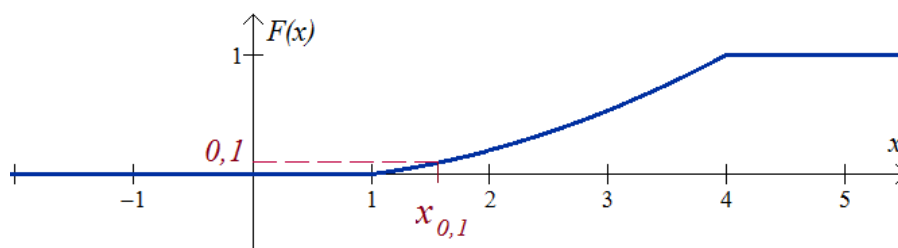
Stručne zapíšeme $F(x) = \begin{cases} 0 & \text{pre } x \in (-\infty, 1); \\ \frac{1}{15} (x^2 - 1) & \text{pre } x \in \langle 1, 4 \rangle; \\ 1 & \text{pre } x \in \langle 4, \infty \rangle. \end{cases}$



Obr. 3.33 Graf distribučnej funkcie spojitej náhodnej premennej X

Vzhľadom na zadanie príkladu, prvý decil určíme pomocou distribučnej funkcie, t.j. hodnotu distribučnej funkcie položíme rovnú $\frac{1}{10}$.

$F(x_{0,1}) = \frac{1}{15}(x_{0,1}^2 - 1) = \frac{1}{10} \Rightarrow x_{0,1} = \pm\sqrt{\frac{15}{10} + 1}$, pričom zo zadania príkladu vyplýva, že $x_{0,1} = -\sqrt{\frac{15}{10} + 1}$ nám nevyhovuje. Prvý decil $x_{0,1} = \sqrt{\frac{15}{10} + 1}$ je zobrazený na nasledujúcom obrázku 3.34.



Obr. 3.34 Kvantil $x_{0,1}$ zobrazený na grafe distribučnej funkcie spojitej náhodnej premennej X

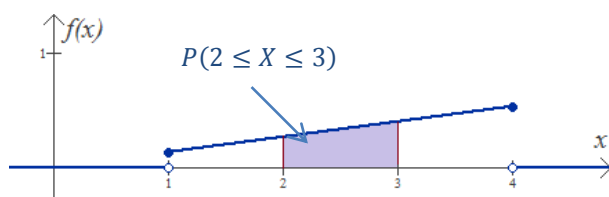
Pravdepodobnosť toho, že náhodná premenná X nadobúda hodnoty od 2 po 3 je možné určiť ako plochu pod grafom funkcie hustoty pravdepodobnosti $f(x)$ medzi 2 a 3 alebo priamo ako rozdiel hodnôt distribučnej funkcie $F(x)$ v $x = 3$ a $x = 2$.

Pravdepodobnosť vyjadrená ako plošný obsah:

$$\begin{aligned} P(2 \leq X \leq 3) &= \int_2^3 \frac{2}{15}x \, dx = \frac{2}{15} \int_2^3 x \, dx = \frac{2}{15} \left[\frac{x^2}{2} \right]_2^3 \\ &= \frac{1}{15} [x^2]_2^3 = \frac{1}{15} \cdot (3^2 - 2^2) = \frac{1}{3}. \end{aligned}$$

Pravdepodobnosť vyjadrená ako rozdiel hodnôt distribučnej funkcie:

$$\begin{aligned} P(2 \leq X \leq 3) &= F(3) - F(2) \\ &= \frac{1}{15}(3^2 - 1) - \frac{1}{15}(2^2 - 1) = \frac{8}{15} - \frac{3}{15} \\ &= \frac{1}{3}. \end{aligned}$$



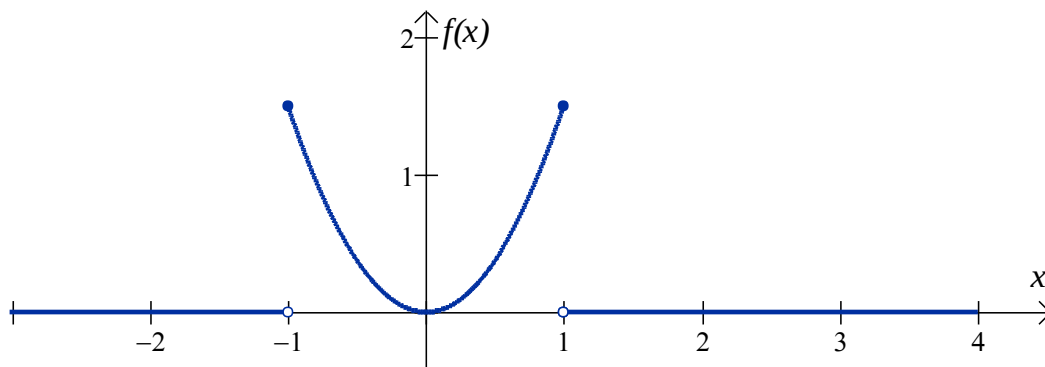
Príklad 3.26 Určte c tak, aby dvojpredpisová funkcia $f(x)$ bola hustotou rozdelenia pravdepodobnosti spojitej náhodnej premennej X : $f(x) = \begin{cases} cx^2 & \text{pre } x \in \langle -1, 1 \rangle; \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle. \end{cases}$

Vypočítajte jej strednú hodnotu a smerodajnú odchýlku.

Riešenie.

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{-1} 0 dx + \int_{-1}^1 cx^2 dx + \int_1^{\infty} 0 dx = 0 + \left[c \cdot \frac{x^3}{3} \right]_{-1}^1 + 0 = \frac{c}{3} \cdot [x^3]_{-1}^1 = \frac{c}{3} \cdot (1^3 - (-1)^3) = \frac{2}{3}c \Rightarrow c = \frac{3}{2}$$

Graf funkcie $f(x) = \begin{cases} \frac{3}{2}x^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}$ môžeme vidieť na nasledujúcom obrázku 3.35.



Obr. 3.35 Graf funkcie hustoty pravdepodobnosti $f(x) = \begin{cases} \frac{3}{2}x^2 & \text{pre } x \in \langle -1, 1 \rangle \\ 0 & \text{pre } x \notin \langle -1, 1 \rangle \end{cases}$

náhodnej premennej X

Stredná hodnota:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_{-\infty}^{-1} x \cdot 0 dx + \int_{-1}^1 x \cdot \frac{3}{2}x^2 dx + \int_1^{\infty} x \cdot 0 dx = \frac{3}{2} \int_{-1}^1 x^3 dx = \frac{3}{2} \cdot \left[\frac{x^4}{4} \right]_{-1}^1 = \frac{3}{8} \cdot [x^4]_{-1}^1 = \\ &= \frac{3}{8} \cdot (1^4 - (-1)^4) = 0. \end{aligned}$$

Smerodajná odchýlka:

Najskôr je potrebné vypočítať rozptyl

$$\begin{aligned} D(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) dx = \int_{-\infty}^{-1} (x - 0)^2 \cdot 0 dx + \int_{-1}^1 (x - 0)^2 \cdot \frac{3}{2}x^2 dx + \int_1^{\infty} (x - 0)^2 \cdot 0 dx = \\ &= \frac{3}{2} \int_{-1}^1 x^4 dx = \frac{3}{2} \cdot \left[\frac{x^5}{5} \right]_{-1}^1 = \frac{3}{10} \cdot [x^5]_{-1}^1 = \frac{3}{10} \cdot (1^5 - (-1)^5) = \frac{3}{5}. \end{aligned}$$

Potom smerodajná odchýlka $s = \sqrt{D(X)} = \sqrt{\frac{3}{5}}$.

3.5 NORMOVANÁ NÁHODNÁ PREMENNÁ

Definícia 3.13 Hovoríme, že náhodná premenná X je normovaná náhodná premenná, ak jej stredná hodnota je nula a disperzia jeden, t.j. $E(X) = 0$ a $D(X) = 1$.

Veta 3.6 (o transformácii normálnej premennej na normovanú)

Nech X je náhodná premenná, pre ktorú existuje $E(X)$ a $D(X) \neq 0$. Potom náhodná premenná $Z = \frac{X-E(X)}{\sqrt{D(X)}} = \frac{X-E(X)}{\sigma(X)}$ je normovaná náhodná premenná.

Poznámka 3.13 Koeficienty šikmosti a špicatosti sa dajú vyjadriť pomocou normovanej náhodnej premennej aj nasledovne $\alpha(X) = E\left(\frac{X-E(X)}{\sigma(X)}\right)^3$; $\beta(X) = E\left(\frac{X-E(X)}{\sigma(X)}\right)^4 - 3$.

Úlohy

3.1 Doplňte chýbajúcu hodnotu pravdepodobnosti p_i , aby tabuľka vyjadrovala zákon rozdelenia pravdepodobnosti náhodnej premennej:

a)

x_i	-1	2	3	4	5	6
p_i	0,14	0,13		0,20	0,11	0,24

b)

x_i	1	2	3	4	5
p_i	$\frac{1}{6}$	$\frac{1}{6}$		$\frac{1}{6}$	$\frac{1}{6}$

c)

x_i	0	1	2
p_i	0,716	0,268	

d)

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

e)

x_i	-3	0	3
p_i	0,3	0,1	

f)

x_i	0	1	2	3
p_i		0,36	0,38	0,14

g)

x_i	-1	0	1	2
p_i		0,1	0,4	0,4

h)

x_i	-2	-1	0	1	2
p_i	0,1	0,2		0,2	0,1

$\left[a) 0,18; b) \frac{2}{6}; c) 0,16; d) 0,2; e) 0,6; f) 0,12; g) 0,1; h) 0,4 \right]$

3.2 Rozdelenie náhodnej premennej X je dané pravdepodobnostnou funkciou

$$p(x) = \begin{cases} c \cdot \left(\frac{6-x}{4}\right) & x = 0, 1, 2, 3 \\ 0 & x \neq 0, 1, 2, 3 \end{cases}$$

Určte hodnotu c a napíšte pravdepodobnostnú tabuľku.

$$\begin{aligned} 1 &= \sum_{x_i} p(x) = \sum_{x=0}^3 c \cdot \left(\frac{6-x}{4}\right) = c \cdot \sum_{x=0}^3 \frac{6-x}{4} = c \cdot \left(\frac{6}{4} + \frac{6-1}{4} + \frac{6-2}{4} + \frac{6-3}{4}\right) = c \cdot \left(\frac{6}{4} + \frac{5}{4} + \frac{4}{4} + \frac{3}{4}\right) \\ &= c \cdot \frac{18}{4} = c \cdot \frac{9}{2} \Rightarrow c = \frac{2}{9} \end{aligned}$$

x_i	0	1	2	3	Σ
p_i	$\frac{2}{9} \cdot \frac{6}{4}$	$\frac{2}{9} \cdot \frac{5}{4}$	$\frac{2}{9} \cdot \frac{4}{4}$	$\frac{2}{9} \cdot \frac{3}{4}$	1

3.3 Nech X je diskretná náhodná premenná daná tabuľkou rozdelenia pravdepodobnosti:

x_i	-1	1	3	5	7
p_i	0,2	0,3	0,2	0,2	0,1

Vypočítajte strednú hodnotu, disperziu, smerodajnú odchýlku, koeficient šikmosti a špicatosti náhodnej premennej X . Vypočítajte pravdepodobnosť $P\left(\frac{2}{3} \leq X < 5\right)$.

$$\left[E(X) = 2,4; D(X) = 6,44; \sigma(X) = 2,54; \alpha(X) = 0,28; \beta(X) = -1,03; P\left(\frac{2}{3} \leq X < 5\right) = 0,5 \right]$$

3.4 V miestnosti je 34 osôb, z nich je 7 žien. Náhodne vyberieme 4 osoby. Nech X je náhodná premenná, ktorá znamená počet žien medzi 4 vybranými. Zostrojte zákon rozdelenia pravdepodobnosti takto definovanej náhodnej premennej X . Vypočítajte kumulatívne pravdepodobnosti, strednú hodnotu a disperziu X .

[Pravdepodobnostná tabuľka:

x_i	0	1	2	3	4	Σ
p_i	0,3784	0,441	0,1589	0,0204	0,00075	1
kumul. p_i	0,3784	0,82	0,9789	0,9992	1	-----

$$E(X) = 0,823529, D(X) = 0,594527]$$

3.5 Nech náhodná premenná X je definovaná ako počet padnutých šestiek pri piatich hodoch hracou kockou. Zostrojte zákon rozdelenia pravdepodobnosti takto definovanej náhodnej premennej X . Vypočítajte kumulatívne pravdepodobnosti, strednú hodnotu a smerodajnú odchýlku X .

[Pravdepodobnostná tabuľka:

xi	0	1	2	3	4	5	Σ
pi	0,4019	0,402	0,1608	0,0322	0,00322	0,00013	1
kumul. pi	0,4019	0,804	0,9645	0,9967	0,99987	1	-----

$$E(X) = 0,83333, \sigma(X) = 0,83333]$$

3.6 V sérii 27 výrobkov je 8 nepodarkov, z ktorých náhodne vyberieme 6 výrobkov. Nech náhodná premenná X znamená počet nepodarkov medzi vybranými. Zostrojte zákon rozdelenia pravdepodobnosti náhodnej premennej X . Vypočítajte kumulatívne pravdepodobnosti, strednú hodnotu X .

[Pravdepodobnostná tabuľka:

xi	0	1	2	3	4	5	6	Σ
pi	0,0917	0,31426	0,3666	0,1833	0,0404	0,0036	9,5E-05	1
kumul. pi	0,0917	0,40592	0,7726	0,9559	0,9963	0,9999	1	-----

$$E(X) = 1,77778]$$

3.7 Medzi kovovými skrutkami sa nachádza 6 so zaoblenou a 4 s plochou hlavou. Náhodne z nich vytiahneme za sebou 5 kusov, pričom každý raz po vytiahnutí skrutku vrátime a premiešame. Náhodnou premennou X je počet vytiahnutých skrutiek so zaoblenou hlavou. Nájdite:

- zákon rozdelenia pravdepodobnosti X ,
- strednú hodnotu,
- disperziu.

[Obor hodnôt náhodnej premennej $X = \{0,1,2,3,4,5\}$, jej pravdepodobnostná tabuľka:

x_i	0	1	2	3	4	5	Σ
p_i	0,01	0,08	0,23	0,35	0,26	0,08	1

$$E(X) = 3,00; D(X) = 1,20]$$

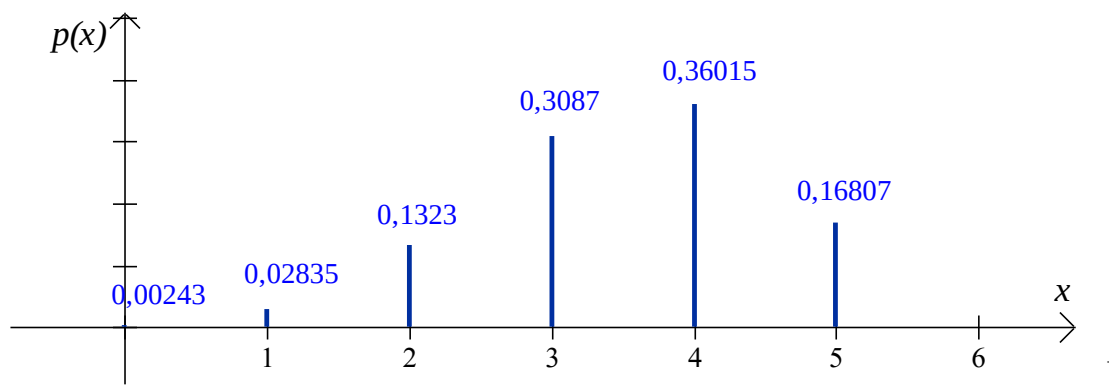
3.8 V dodanej zásielke sa nachádza 7 predlohových a 3 planétové prevodovky. Náhodne z nich vytiahneme za sebou 5 kusov, pričom každý kus po vytiahnutí vrátime a následne všetky premiešame. Vyjadrite pomocou Bernoulliho schémy pravdepodobnosť toho, že medzi náhodne vybranými kusmi nebude ani jedna predlohová prevodovka a pravdepodobnosť toho, že vytiahneme iba jednu predlohovú prevodovku. Nech je náhodnou premennou X počet vytiahnutých predlohových prevodoviek. Nájdite zákon rozdelenia pravdepodobnosti X . Nakreslite paličkový graf.

$$\left[p(X=0) = \binom{5}{0} \cdot 0,7^0 \cdot 0,3^5 = 0,00243; p(X=1) = \binom{5}{1} \cdot 0,7^1 \cdot 0,3^{5-1} = 0,02835; \text{ obor hodnôt} \right.$$

náhodnej premennej $X = \{0,1,2,3,4,5\}$; jav A – vytiahnutie predlohovej prevodovky pri jednom (každom) pokuse; $P(A) = \frac{7}{10} = 0,7$; pravdepodobnostná tabuľka:

x_i	0	1	2	3	4	5	Σ
p_i	0,00243	0,02835	0,1323	0,3087	0,36015	0,16807	1

Paličkový graf:



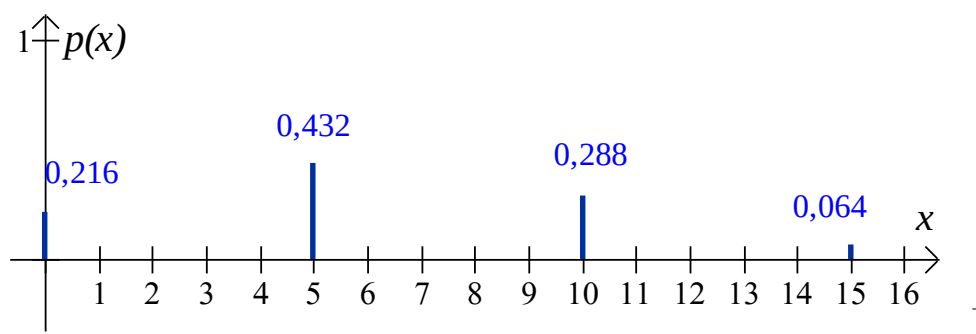
3.9 Strelec vystrelí do terča trikrát. Pravdepodobnosť zásahu pri každom výstrele je 0,4. Za každý zásah dostane strelec 5 bodov. Určte, aké hodnoty môže náhodná premenná X , vyjadrujúca počet získaných bodov, nadobudnúť a vypočítajte k nim prislúchajúce pravdepodobnosti. Zostrojte tabuľku rozdelenia pravdepodobnosti počtu získaných bodov. Zobrazte graficky.

[$X = \{0 \text{ zásahov} \rightarrow 0 \text{ bodov}, 1 \text{ zásah} \rightarrow 5 \text{ bodov}, 2 \text{ zásahy} \rightarrow 10 \text{ bodov}, 3 \text{ zásahy} \rightarrow 15 \text{ bodov}\}$]

$$p(X=0) = \binom{3}{0} \cdot 0,4^0 \cdot 0,6^3 = 0,216; \quad p(X=5) = \binom{3}{1} \cdot 0,4^1 \cdot 0,6^{3-1} = 0,432;$$

$$p(X=10) = \binom{3}{2} \cdot 0,4^2 \cdot 0,6^{3-2} = 0,288; \quad p(X=15) = \binom{3}{3} \cdot 0,4^3 \cdot 0,6^{3-3} = 0,064;$$

x_i	0	5	10	15	Σ
p_i	0,216	0,432	0,288	0,064	1



3.10 Náhodná premenná X nadobúda hodnoty i s pravdepodobnosťou $p_i = P(X = i) = \frac{1}{6}$,

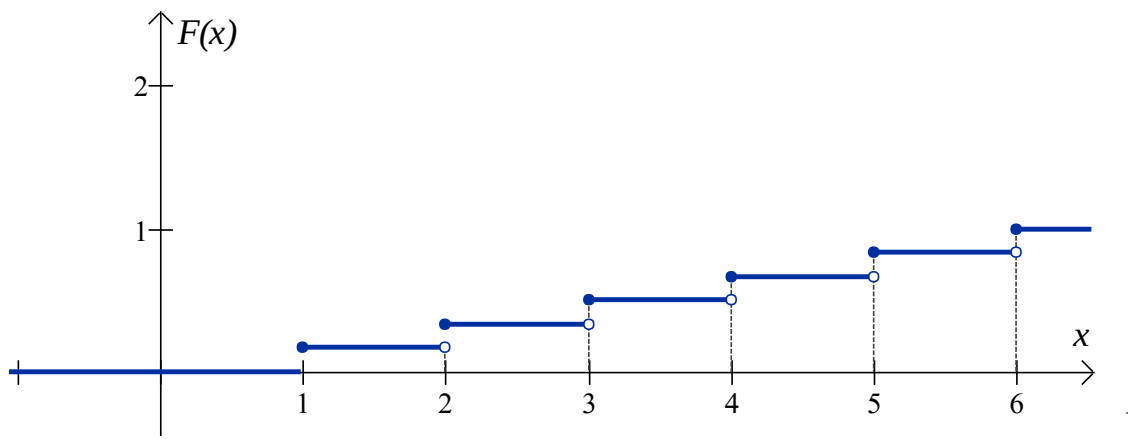
kde $i = 1, 2, \dots, 6$. Zostrojte distribučnú funkciu a nakreslite jej graf.

[

x_i	1	2	3	4	5	6	Σ
p_i	1/6	1/6	1/6	1/6	1/6	1/6	1

Pre distribučnú funkciu $F(x) = \sum_{x_i \leq x} P(X = x_i)$ pre $-\infty < x < \infty$ platí:

$$\begin{aligned}
 x < 1: \quad F(x) &= 0 \\
 1 \leq x < 2: \quad F(x) &= \frac{1}{6} \\
 2 \leq x < 3: \quad F(x) &= \frac{2}{6} = \frac{1}{3} \\
 3 \leq x < 4: \quad F(x) &= \frac{3}{6} = \frac{1}{2} \\
 4 \leq x < 5: \quad F(x) &= \frac{4}{6} = \frac{2}{3} \\
 5 \leq x < 6: \quad F(x) &= \frac{5}{6} \\
 x \geq 6: \quad F(x) &= \frac{6}{6} = 1
 \end{aligned}
 \Rightarrow F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < 1; \\ \frac{1}{6} & \text{pre } 1 \leq x < 2; \\ \frac{1}{3} & \text{pre } 2 \leq x < 3; \\ \frac{1}{2} & \text{pre } 3 \leq x < 4; \\ \frac{2}{3} & \text{pre } 4 \leq x < 5; \\ \frac{5}{6} & \text{pre } 5 \leq x < 6; \\ 1 & \text{pre } x \geq 6; \end{cases}$$



3.11 Nech náhodná premenná X nadobúda hodnoty 0, 1 s pravdepodobnosťou $P(X = 0) = p$, $P(X = 1) = 1 - p$. Určte distribučnú funkciu a znázornite graficky.

[

x_i	0	1	Σ
p_i	p	$1 - p$	1

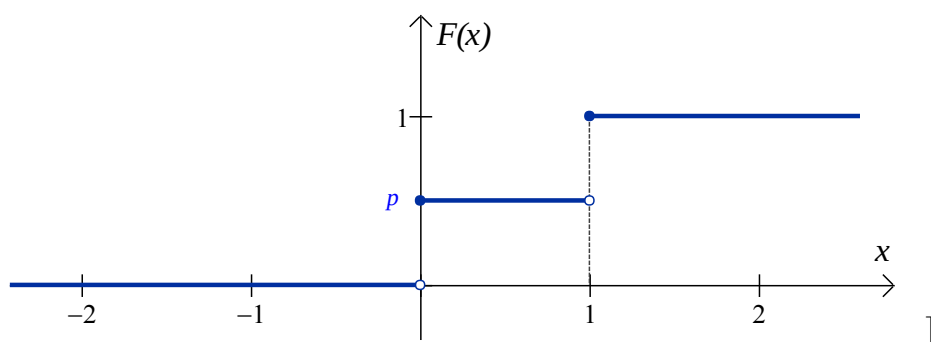
Distribučná funkcia $F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$ pre $-\infty < x < \infty$

$$x < 0: \quad F(x) = 0$$

$$0 \leq x < 1: \quad F(x) = P(X \leq x) = P(X = 0) = p$$

$$x \geq 1: \quad F(x) = P(X \leq x) = P(X = 0) + P(X = 1) = p + 1 - p = 1$$

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < 0; \\ p & \text{pre } 0 \leq x < 1; \\ 1 & \text{pre } x \geq 1; \end{cases}$$

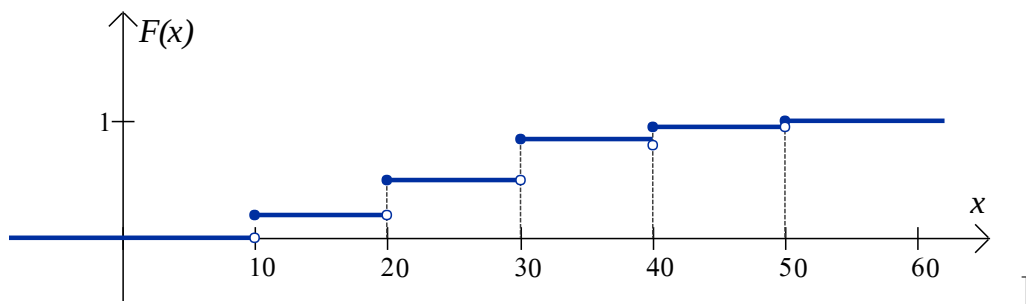


3.12 Zákon rozdelenia diskkrétnej náhodnej premennej X je daný tabuľkou:

x_i	10	20	30	40	50	Σ
p_i	0,2	0,3	0,35	0,1	0,05	1

Určte distribučnú funkciu náhodnej premennej X . Znázornite graficky.

$$[\quad F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < 10; \\ 0,2 & \text{pre } 10 \leq x < 20; \\ 0,5 & \text{pre } 20 \leq x < 30; \\ 0,85 & \text{pre } 30 \leq x < 40; \\ 0,95 & \text{pre } 40 \leq x < 50; \\ 1 & \text{pre } x \geq 50; \end{cases}$$

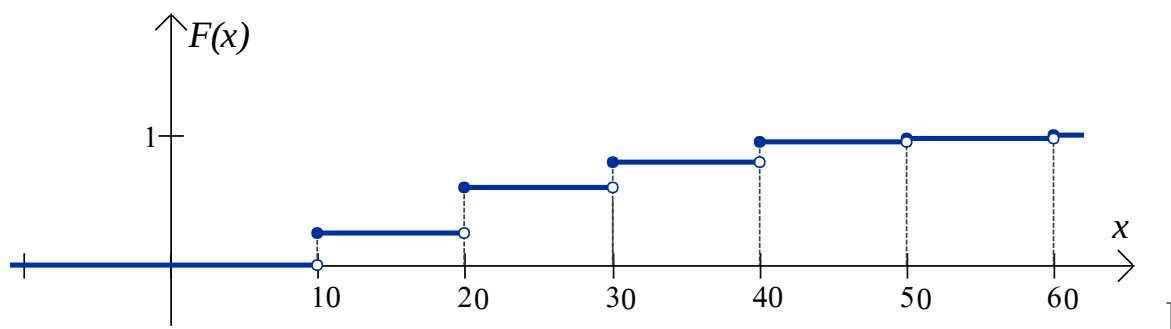


3.13 Zákon rozdelenia diskkrétnej náhodnej premennej X je daný tabuľkou:

x_i	10	20	30	40	50	60
p_i	0,24	0,36	0,2	0,15	0,03	0,02

Určte distribučnú funkciu náhodnej premennej X . Znázornite graficky.

$$[\quad F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < 10; \\ 0,24 & \text{pre } 10 \leq x < 20; \\ 0,6 & \text{pre } 20 \leq x < 30; \\ 0,8 & \text{pre } 30 \leq x < 40; \\ 0,95 & \text{pre } 40 \leq x < 50; \\ 0,98 & \text{pre } 50 \leq x < 60; \\ 1 & \text{pre } x \geq 60; \end{cases}$$



3.14 Zákon rozdelenia diskkrétnej náhodnej premennej X je daný tabuľkou:

x_i	2	4	8	10
p_i	0,4	0,2	0,1	0,3

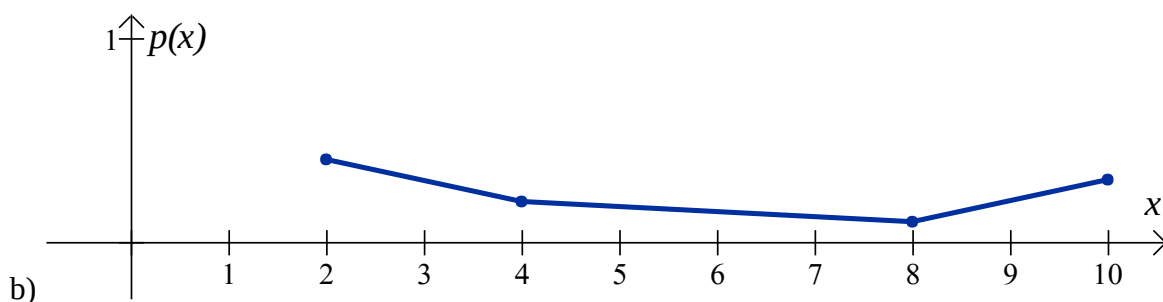
Nájdite:

a) $P(2 < X \leq 4)$;

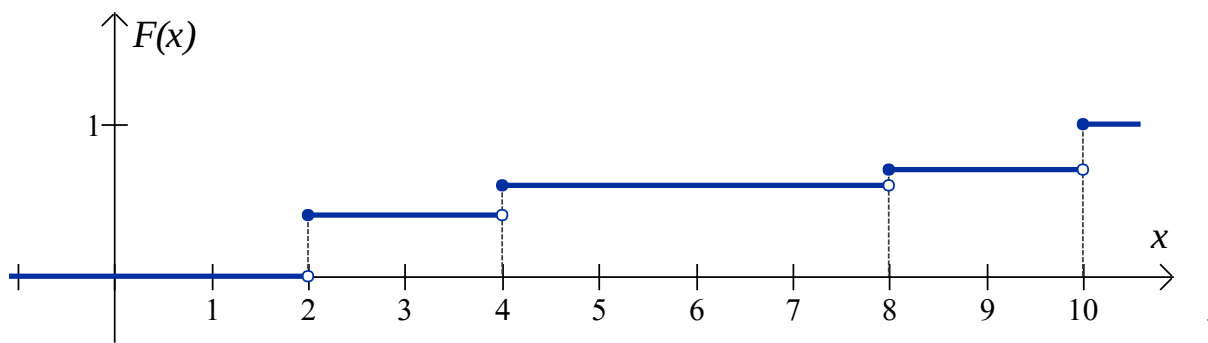
b) polygón rozdelenia;

c) distribučnú funkciu náhodnej premennej X . Znázornite graficky.

[a] $P(2 < X \leq 4) = P(X = 4) = 0,2$;



$$c) F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < 2; \\ 0,4 & \text{pre } 2 \leq x < 4; \\ 0,6 & \text{pre } 4 \leq x < 8; \\ 0,7 & \text{pre } 8 \leq x < 10; \\ 1 & \text{pre } x \geq 10; \end{cases}$$



3.15 Zákón rozdelenia pravdepodobnosti náhodnej premennej X je daný tabuľkou:

x_i	0	1	2	3	4
p_i	0,2	0,4	0,3	0,08	0,02

Nájdite jej strednú hodnotu, disperziu, smerodajnú odchýlku, modus.

$$\left[E(X) = 1,32; D(X) = E(X^2) - [E(X)]^2 = 2,64 - 1,32^2 = 0,8976; \sigma(X) = \sqrt{D(X)} = \sqrt{0,8976} = 0,9474; Mo = 1 \right]$$

3.16 Rozdelenie diskkrétnej náhodnej premennej X je dané matematickou formulou (pravdepodobnostnou funkciou):

$$p(x) = \begin{cases} c \cdot \frac{x}{2} & x = 2, 4, 6, 8; \\ 0 & x \neq 2, 4, 6, 8. \end{cases}$$

Nájdite:

- konštantu c ,
- distribučnú funkciu náhodnej premennej X ,
- vypočítajte $E(X)$, $D(X)$, $\sigma(X)$, Mo .

[a)

x_i	2	4	6	8	Σ
p_i	c	$2c$	$3c$	$4c$	1
p_i	0,1	0,2	0,3	0,4	

$$\sum_{x_i} P(X = x_i) = 1 \Rightarrow c + 2c + 3c + 4c = 1 \Rightarrow c = \frac{1}{10}$$

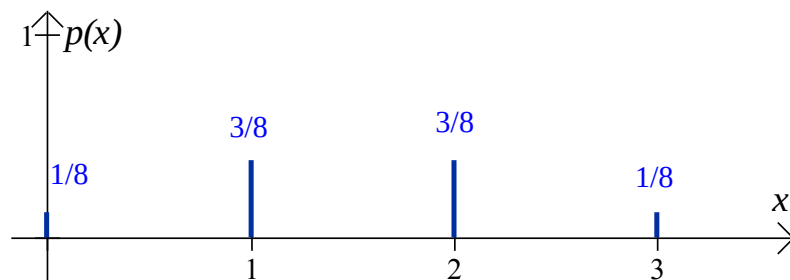
$$b) F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < 2; \\ 0,1 & \text{pre } 2 \leq x < 4; \\ 0,3 & \text{pre } 4 \leq x < 6; \\ 0,6 & \text{pre } 6 \leq x < 8; \\ 1 & \text{pre } x \geq 8; \end{cases}$$

$$c) E(X) = 6; D(X) = E(X^2) - [E(X)]^2 = 40 - 6^2 = 4 \Rightarrow \sigma(X) = \sqrt{D(X)} = \sqrt{4} = 2; Mo = 8]$$

3.17 V tabuľke doplňte chýbajúce údaje tak, aby bola zákonom rozdelenia pravdepodobnosti nejakej diskkrétnej náhodnej premennej X . Načrtnite paličkový graf rozdelenia náhodnej premennej X . Vypočítajte Mo , $E(X)$, $D(X)$, $\sigma(X)$, $\alpha(X)$, $\beta(X)$ a prvé štyri začiatkové a centrálné momenty náhodnej premennej X . Nájdite distribučnú funkciu náhodnej premennej X a načrtnite jej graf.

x_i	0	1	2	3	Σ
p_i	1/8	3/8		1/8	1

$$[1 = \sum_i p_i \Rightarrow p(X=2) = \frac{3}{8}; \text{paličkový graf:}$$



$Mo = 1, Mo = 2 \Rightarrow$ náhodná premenná X má bimodálne rozdelenie;

začiatkové momenty:

$$z_1 = E(X) = \sum_i x_i \cdot p_i = \frac{12}{8} = \frac{3}{2} = 1,5; z_2 = E(X^2) = \sum_i x_i^2 \cdot p_i = \frac{24}{8} = 3;$$

$$z_3 = E(X^3) = \sum_i x_i^3 \cdot p_i = \frac{54}{8} = 13,5; z_4 = E(X^4) = \sum_i x_i^4 \cdot p_i = \frac{132}{8} = 16,5;$$

centrálné momenty:

$$c_1 = \sum_i (x_i - \bar{x}) \cdot p_i = 0;$$

$$c_2 = \sum_i (x_i - \bar{x})^2 \cdot p_i = D(x) = E(X^2) - [E(X)]^2 = z_2 - z_1^2 = 3 - 2,25 = 0,75;$$

$$\sigma(X) = \sqrt{D(x)} = \sqrt{0,75} = 0,866;$$

$$c_3 = z_3 - 3z_2 z_1 + 2z_1^3 = 13,5 - 3 \cdot 3 \cdot 1,5 + 2 \cdot (1,5)^3 = 6,75;$$

$$c_4 = z_4 - 4z_3 z_1 + 6z_2 z_1^2 - 3z_1^4 = 16,5 - 4 \cdot 13,5 \cdot 1,5 + 6 \cdot 3 \cdot (1,5)^3 = -3,75.$$

Distribučná funkcia:

$$x < 0: \quad F(x) = 0$$

$$0 \leq x < 1: \quad F(x) = P(X \leq x) = P(X = 0) = \frac{1}{8}$$

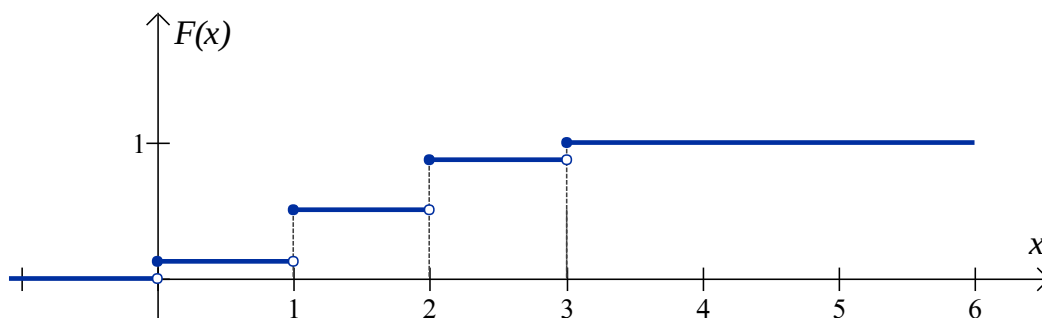
$$1 \leq x < 2: \quad F(x) = P(X \leq x) = P(X = 0) + P(X = 1) = \frac{4}{8}$$

$$2 \leq x < 3: \quad F(x) = P(X \leq x) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{7}{8}$$

$$x \geq 3: \quad F(x) = P(X \leq x) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$$

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < 0; \\ \frac{1}{8} & \text{pre } 0 \leq x < 1; \\ \frac{4}{8} & \text{pre } 1 \leq x < 2; \\ \frac{7}{8} & \text{pre } 2 \leq x < 3; \\ 1 & \text{pre } x \geq 3; \end{cases}$$

Graf distribučnej funkcie:



3.18 V zásielke sa nachádza 20 výrobkov, z nich 4 sú nepodarky. Náhodne vyberieme 2 výrobky. Náhodnou premennou X je počet nepodarkov. Napíšte pravdepodobnostnú tabuľku doplnenú o kumulatívne postupnosti.

x_i	0	1	2	Σ
p_i	0,63	0,34	0,03	1
$\sum_i p_i$	0,63	0,97	1	

3.19 V zásielke 50 výrobkov je 5 nepodarkov. Náhodne vyberieme 4 výrobky. Náhodnou premennou X je počet nepodarkov medzi vybranými výrobkami. Určte hodnoty, ktoré X nadobúda a vypočítajte pravdepodobnosti, ktoré im prislúchajú.

$$[X = \{0,1,2,3,4\}; P(X=0) = \frac{C_4(45)}{C_4(50)} = \frac{148995}{230300} = 0,64;$$

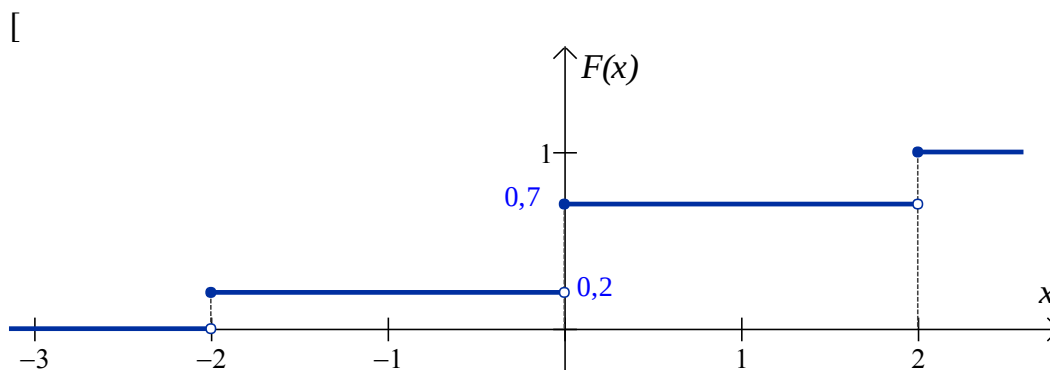
$$P(X=1) = \frac{5 \cdot C_3(45)}{C_4(50)} = \frac{70950}{230300} = 0,31608; P(X=2) = \frac{C_2(5) \cdot C_2(45)}{C_4(50)} = \frac{10 \cdot 990}{230300} = 0,042;$$

$$P(X=3) = \frac{C_3(5) \cdot C_1(45)}{C_4(50)} = \frac{50 \cdot 45}{230300} = 0,0019; P(X=4) = \frac{C_4(5) \cdot C_0(45)}{C_4(50)} = \frac{5}{230300} = 0,00002]$$

3.20 Nech diskretná náhodná premenná X má distribučnú funkciu

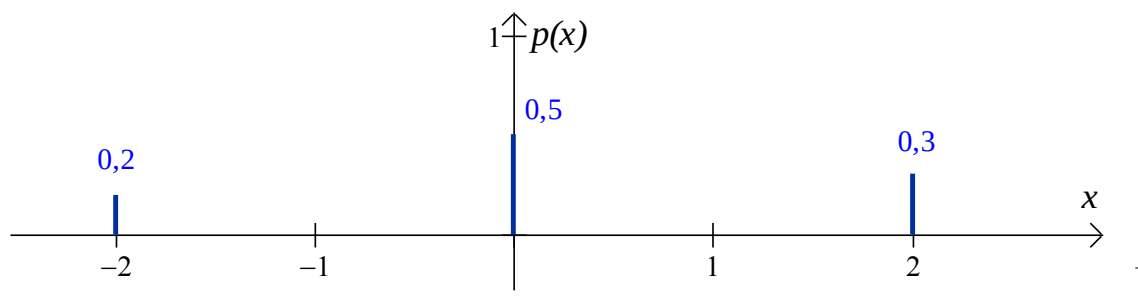
$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < -2; \\ 0,2 & \text{pre } -2 \leq x < 0; \\ 0,7 & \text{pre } 0 \leq x < 2; \\ 1 & \text{pre } x \geq 2. \end{cases}$$

Zobrazte jej graf. Zostrojte pravdepodobnostnú tabuľku a pomocou paličkového grafu zobrazte jej pravdepodobnostnú funkciu.



Skoky v bodoch x_i v distribučnej funkcii sú rovné hodnote $p(x_i) = P(X = x_i)$.

x_i	-2	0	2	Σ
p_i	0,2	0,5	0,3	1
$\sum_i p_i$	0,2	0,7	1	-



3.21 V urne sa nachádza 5 bielych a 25 čiernych gulôčok. Náhodne vyberieme jednu gulôčku. Náhodnou premennou X je vytiahnutie bielej gulôčky. Nájdite distribučnú funkciu náhodnej premennej X .

[

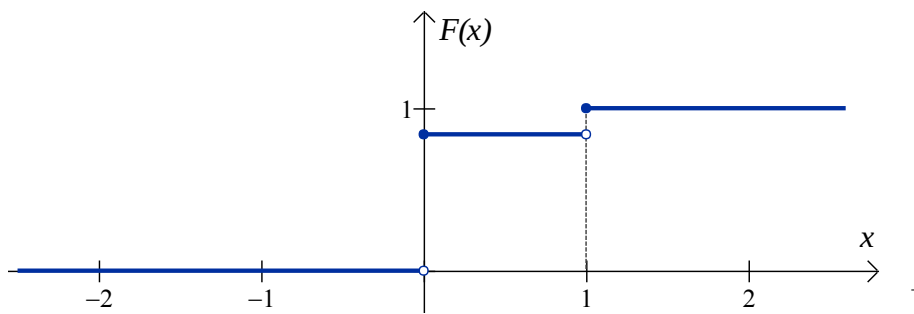
x_i	0	1
p_i	$5/6$	$1/6$

$$x < 0: \quad F(x) = 0$$

$$0 \leq x < 1: \quad F(x) = \frac{5}{6}$$

$$x \geq 1: \quad F(x) = \frac{5}{6} + \frac{1}{6} = 1$$

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x < 0; \\ \frac{5}{6} & \text{pre } 0 \leq x < 1; \\ 1 & \text{pre } x \geq 1. \end{cases}$$



4. ROZDELENIA NÁHODNEJ PREMENNEJ

Diskrétna náhodná premenná je zadaná, ak poznáme všetky jej možné hodnoty $x_1, x_2, \dots, x_n, \dots$ a pravdepodobnosti $P(X = x_i) = p_i$, pričom $\sum p_i = 1$, t.j. je známa jej pravdepodobnostná funkcia.

Spojité náhodná premenná je zadaná, ak poznáme jej hustotu rozdelenia pravdepodobnosti.

Pomocou pravdepodobnostnej funkcie diskkrétnej náhodnej premennej či funkcie hustoty spojitkej náhodnej premennej sú popísané rôzne pravdepodobnostné rozdelenia náhodných premenných.

4.1 NIEKTORÉ ROZDELENIA DISKRÉTNEJ NÁHODNEJ PREMENNEJ

4.1.1 Rovnomerné diskkrétne rozdelenie (Uniform discrete distribution)

Ak náhodná premenná X môže nadobúdať hodnoty $1, 2, 3, \dots, n$, každú s rovnakou pravdepodobnosťou výskytu $\frac{1}{n}$, hovoríme o rovnomernom rozdelení premennej. Prirodzené číslo n je jediný parameter rozdelenia, čo všeobecne zapisujeme $X \sim R(n)$.

Pravdepodobnostná tabuľka:

x_i	1	2	3		n	$\sum p_i$
p_i	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$		$\frac{1}{n}$	1

čo môžeme zapísať aj predpisom

$X \sim R(n)$	
$p(x)$	$= \begin{cases} \frac{1}{n} & \text{pre } x_1, x_2, \dots, x_n; \\ 0 & \text{inak} \end{cases}$

Číselné charakteristiky náhodnej premennej s rovnomerným diskrétnym rozdelením sú:

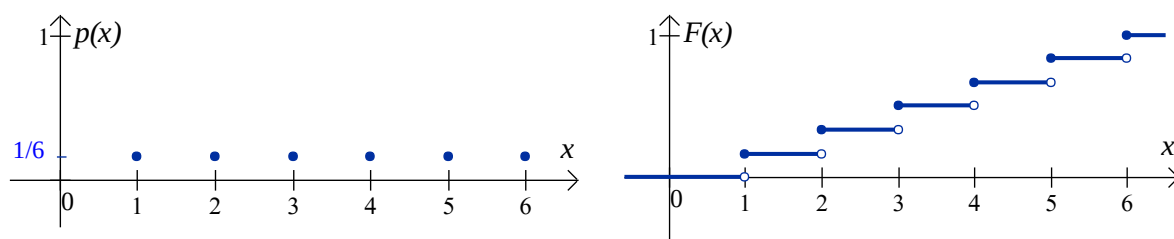
$E(X)$	$D(X)$	$\alpha(X)$	$\beta(X)$
$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	0	$-\frac{6(n^2+1)}{5(n^2-1)}$

Príklad 4.1 Typickým príkladom je náhodná premenná, popisujúca počet padnutých bodov na hracej kocke. Výsledkom po hode kockou je jeden zo 6 možných javov, každý nastane

s pravdepodobnosťou $1/6$, t.j. pravdepodobnostná funkcia je daná vzťahom $p(x_i) = 1/6$. Jej stredná hodnota $E(X) = \frac{7}{2}$.

x_i	1	2	3	4	5	6	Σ
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1
$x_i \cdot p_i$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	$\frac{6}{6}$	$\frac{21}{6} = \frac{7}{2}$

Graf pravdepodobnostnej a distribučnej funkcie je na obr. 4.1.



Obr. 4.1 Pravdepodobnostná a distribučná funkcia rozdelenia $R(6)$

Úlohy

4.1 Posledná číslica sériového čísla súčiastky je s rovnakou pravdepodobnosťou jedna z čísiel 0 až 9. Náhodný experiment spočíva vo výbere jednej súčiastky z veľkej dávky, pričom nás zaujíma prvá číslica sériového čísla súčiastky. Určte pravdepodobnostnú a distribučnú funkciu náhodnej premennej X - posledná číslica sériového čísla vybranej súčiastky.

$$p(x) = \begin{cases} 0,1 & \text{pre } x = 0, 1, \dots, 9 \\ 0 & \text{inak} \end{cases}; F(x) = 0 \text{ pre } x \in (-\infty, 0), F(x) = 0,1 \text{ pre } x \in (0,1), F(x) = 0,2 \text{ pre } x \in (1,2), F(x) = 0,3 \text{ pre } x \in (2,3), F(x) = 0,4 \text{ pre } x \in (3,4), F(x) = 0,5 \text{ pre } x \in (4,5), F(x) = 0,6 \text{ pre } x \in (5,6), F(x) = 0,7 \text{ pre } x \in (6,7), F(x) = 0,8 \text{ pre } x \in (7,8), F(x) = 0,9 \text{ pre } x \in (8,9), F(x) = 1 \text{ pre } x \in (9, \infty)$$

4.1.2 Alternatívne rozdelenie

Niektoré náhodné pokusy môžu mať iba dva rôzne výsledky: pokus je úspešný a pokus je neúspešný. Náhodná premenná X , udávajúca počet úspechov v jednom pokuse, sa nazýva alternatívna, označujeme $X \sim A(\pi)$, kde parameter rozdelenia π vyjadruje pravdepodobnosť úspechu. Náhodná premenná X priradí číslo 1, ak jav nastane a číslo 0, ak jav nenastane. Podľa toho, že nadobúda hodnoty 0 alebo 1, nazýva sa nula – jednotkové rozdelenie.

x_i	0	1	Σ
p_i	$1 - \pi$	π	1

Zákon rozdelenia pravdepodobnosti (pravdepodobnostná funkcia):

$X \sim A(\pi)$	
$p(x)$	$= \begin{cases} \pi^x(1-\pi)^{1-x} & \text{pre } x = 0, 1, \\ 0 & \text{inak} \end{cases}$

Jej číselné charakteristiky sú:

$E(X)$	$D(X)$	$\alpha(X)$	$\beta(X)$
π	$\pi(1-\pi)$	$\frac{1-2\pi}{\sqrt{\pi(1-\pi)}}$	$\frac{1-6\pi(1-\pi)}{\pi(1-\pi)}$

Príklady alternatívneho rozdelenia pravdepodobnosti :

- počet nepodarkov pri náhodnom výbere jedného výrobku;
- počet zásahov pri jednom výstrele;
- počet spojení pri jednom vytočení telefónneho čísla;

(všeobecne indikuje nastúpenie alebo nenastúpenie sledovaného náhodného javu).

Príklad 4.2 Z dodávky obsahujúcej 15% nekvalitných výrobkov odoberieme jeden. Aká je pravdepodobnosť, že to nebude nekvalitný výrobok?

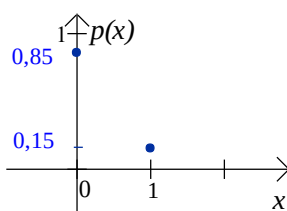
Riešenie. jav B – vyberieme nekvalitný výrobok;

jav $\bar{B}$ – vyberieme kvalitný výrobok;

Pravdepodobnosť, že vyberieme nekvalitný výrobok, t.j. že nastane jav B je $p(1) = 0,15$.

Pravdepodobnosť opačného javu je $p(0) = 1 - 0,15 = 0,85$.

Ak náhodná premenná X predstavuje počet vytiahnutých nekvalitných výrobkov pri jednom výbere, potom platí $X \sim A(0,15)$. Graf pravdepodobnostnej funkcie $p(x)$ takto definovanej náhodnej premennej X je na obr. 4.2.



Obr. 4.2 Pravdepodobnostná funkcia rozdelenia $A(0,15)$

Pravdepodobnosť toho, že pri náhodnom výbere nevyberieme nekvalitný výrobok je 85%.

4.1.3 Binomické rozdelenie (Binomial distribution)

Binomické rozdelenie sa používa pri výbere s opakovaním, t.j. pri sérii n nezávislých opakovaných pokusoch, realizovaných za rovnakých podmienok. Pri každom z nich môže sledovaný jav A nastať s rovnakou pravdepodobnosťou π a nenastať s pravdepodobnosťou $1 -$

π (tzv. Bernoulliho pokus). Pravdepodobnosť javu A pri ľubovoľnom pokuse nezávisí od výsledku predchádzajúcich pokusov. Náhodná premenná X , ktorá udáva koľkokrát jav A pri n -násobnom opakovaní pokusu nastal, môže nadobúdať niektorú hodnotu z množiny $\{0, 1, 2, \dots, n\}$, t.j. spolu $n + 1$ rôznych hodnôt. Rozdelenie pravdepodobnosti tejto náhodnej premennej X sa nazýva binomické rozdelenie.

$X \sim \text{Bi}(n, \pi)$	
$p(x) = \begin{cases} \binom{n}{x} \pi^x (1 - \pi)^{n-x} & \text{pre } x = 0, 1, \dots, n, \\ 0 & \text{inak} \end{cases}$	
Excel:	=BINOM.DIST(číslo_s; pokusy; pravdepodobnosť_s; kumulatívne) číslo_s je hodnota náhodnej premennej X , pokusy sú počet nezávislých pokusov (parameter n), pravdepodobnosť_s je pravdepodobnosť úspechu v jednom pokuse (parameter π), kumulatívne je logická premenná určujúca, či bude vypočítaná pravdepodobnostná funkcia (0) alebo distribučnú funkcia (1).

Poznámka 4.1 Zvláštnym prípadom binomického rozdelenia pre $n = 1$ je alternatívne rozdelenie.

Číselné charakteristiky:

$E(X)$	$D(X)$	$\alpha(X)$	$\beta(X)$
$n\pi$	$n\pi(1 - \pi)$	$\frac{1 - 2\pi}{\sqrt{n\pi(1 - \pi)}}$	$\frac{1 - 6\pi(1 - \pi)}{n\pi(1 - \pi)}$

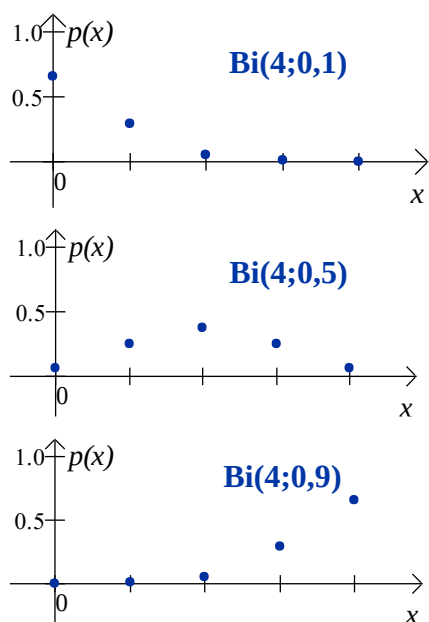
Príklady binomického rozdelenia:

V praxi sa s binomickým rozdelením stretneme pri skúmaní kvality produkcie, keď nás zaujíma, aká časť výrobkov z celkového množstva bude dobrá, resp. bude mať kaz.

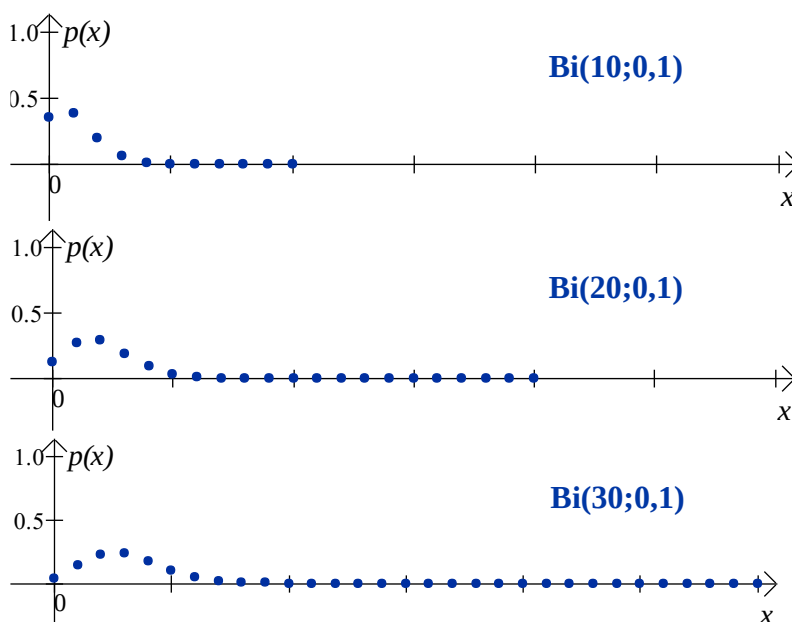
Náhodná premenná X má binomické rozdelenie, ak jej hodnoty predstavujú napr.:

- počet chybných výrobkov zistených vo výbere n výrobkov, vyrábaných určitým strojom za stálych podmienok;
- počet chybných výrobkov zistených vo výbere n výrobkov z danej dávky, pri ktorom výrobok do nej vraciame;
- počet prípadov, v ktorých sa prejavila účinnosť podaného prípravku skúšaného na n objektoch;
- počet prípadov, v ktorých doba výrobnnej operácie prekročila určenú hodnotu, za predpokladu, že bola evidovaná doba n výrobných operácií;
- počet spojení pri n volaniach;
- počet narodených dievčat medzi n novorodencami;

Nasledujúce obrázky 4.3 a 4.4 ukazujú, ako sa mení tvar rozdelenia v závislosti od parametrov π a n .

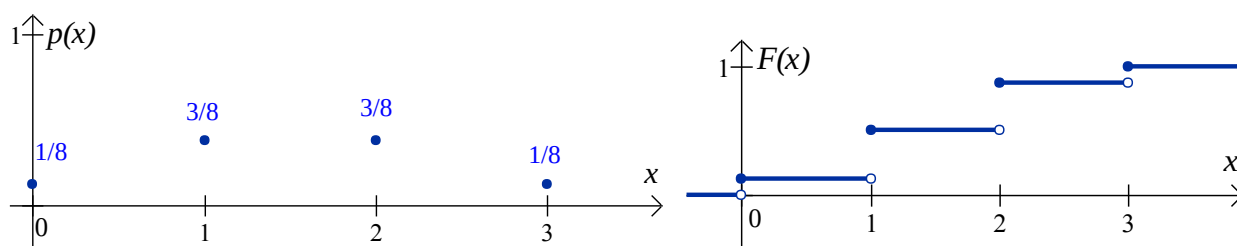


Obr. 4.3 Parameter p ovplyvňuje šikmosť rozdelenia



Obr. 4.4 S rastúcim n sa rozdelenie stáva stále viac symetrickým a konverguje k normálnemu rozdeleniu

Príklad 4.3 Pravdepodobnostná a distribučná funkcia náhodnej premennej X , udávajúcej počet chlapcov medzi 3 po sebe narodenými deťmi pri predpoklade, že pravdepodobnosť narodenia chlapca i dievčaťa sú rovnaké, sú na nasledujúcom obr. 4.5.



Obr. 4.5 Pravdepodobnostná a distribučná funkcia rozdelenia $Bi(3; 0,5)$

Príklad 4.4 Šiestim pacientom bol podaný liek, ktorého úspešnosť je 85%. Určte pravdepodobnosť toho, že liek pomôže:

- práve štyrom pacientom,
- viac ako štyrom pacientom,
- menej ako štyrom pacientom.

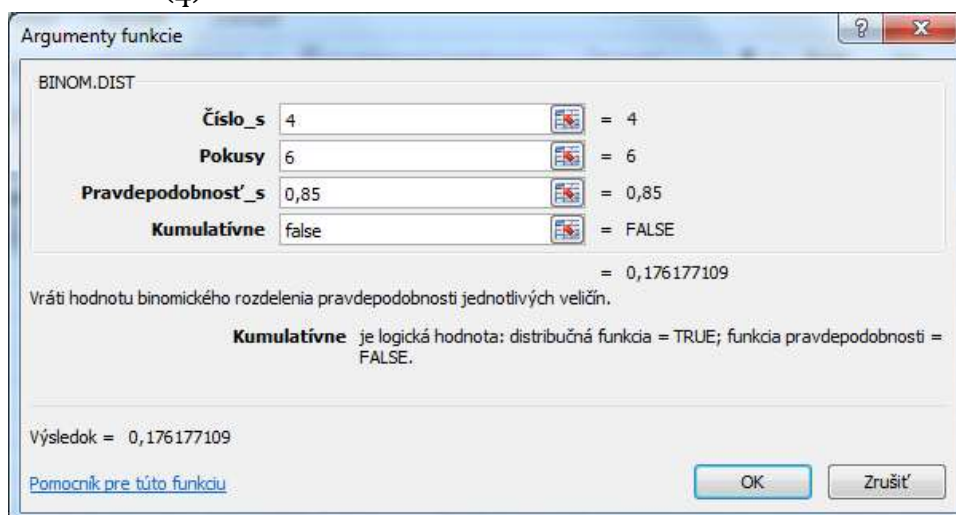
Riešenie. Podanie lieku šiestim pacientom možno považovať za nezávislé 6 – násobné opakovanie náhodného pokusu s konštantnou pravdepodobnosťou $\pi = 0,85$. Uvažujeme náhodnú premennú X – počet pacientov, ktorým podávaný liek pomôže, ktorá má binomické rozdelenie $Bi(6; 0,85)$.

Pravdepodobnostnú funkciu možno popísať pomocou Bernoulliho schémy

$P_6(k) = \binom{6}{k} \cdot 0,85^k \cdot (1 - 0,85)^{6-k}$, pre $k = 0, 1, 2, 3, 4, 5, 6$ je počet pacientov, ktorým liek pomôže.

a) Pravdepodobnosť toho, že zo šiestich pacientov liek pomôže práve štyrom, je preto rovná

$$P(k = 4) = P_6(4) = \binom{6}{4} \cdot 0,85^4 \cdot (1 - 0,85)^{6-4} = 15 \cdot 0,522 \cdot 0,0225 = 0,1762.$$



Obr. 4.6 Výpočet realizovaný v Exceli pomocou funkcie BINOM.DIST

Pravdepodobnosť toho, že liek, ktorý podáme šiestim pacientom, pomôže práve štyrom z nich, je 17,62%.

b) Pravdepodobnosť toho, že zo šiestich pacientov liek pomôže viac ako štyrom je rovná

$$P(k > 4) = P_6(k = 5) + P_6(k = 6) = P_6(5) + P_6(6) = \binom{6}{5} \cdot 0,85^5 \cdot (1 - 0,85)^{6-5} + \binom{6}{6} \cdot 0,85^6 \cdot (1 - 0,85)^{6-6} = 6 \cdot 0,444 \cdot 0,15 + 1 \cdot 0,377 \cdot 1 = 0,399 + 0,377 = 0,7761.$$

	A	
1	0,399335	=BINOM.DIST(5;6;0,85;FALSE)
2	0,37715	=BINOM.DIST(6;6;0,85;FALSE)
3	0,776484	=SUM(A1:A2)
4		

Obr. 4.7 Výpočet realizovaný pomocou Excelu

	A	B	C
1	=BINOM.DIST(5;6;0,85;FALSE)		
2	=BINOM.DIST(6;6;0,85;FALSE)		
3	=SUM(A1:A2)		

Obr. 4.8 Zápis použitých formul

Pravdepodobnosť toho, že zo šiestich pacientov pomôže liek viac ako štyrom, sme vyjadrili ako súčet hodnôt pravdepodobnostnej funkcie pre $k = 5$ a $k = 6$. Hľadaná pravdepodobnosť je 77,61%.

c) Na výpočet pravdepodobnosti toho, že zo šiestich pacientov liek pomôže menej ako štyrom, môžeme použiť distribučnú funkciu – hľadáme jej hodnotu v $k = 3$.

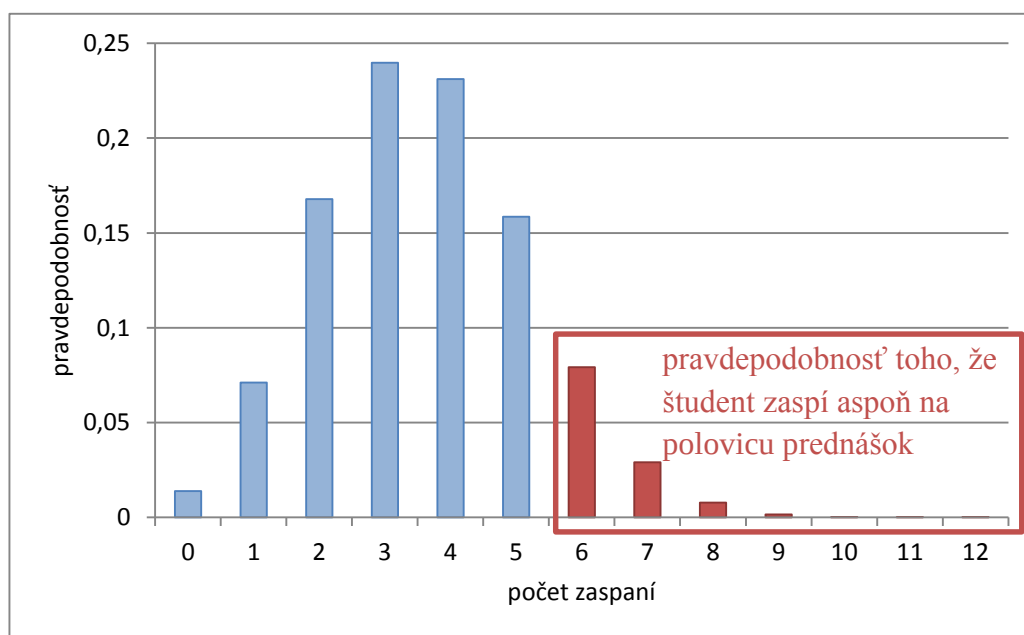
$$P(k < 4) = P(k = 0 \vee k = 1 \vee k = 2 \vee k = 3) = P_6(k = 0) + P_6(k = 1) + P_6(k = 2) + P_6(k = 3) = \text{BINOM.DIST}(3; 6; 0,85; 1) = 0,0477$$

Pravdepodobnosť toho, že zo šiestich pacientov liek pomôže menej ako štyrom (t.j. najviac trom pacientom) je 4,7%.

Príklad 4.5 Študent, ktorý má problémy s ranným vstávaním, zaspí v priemere na 3 prednášky z desiatich ranných prednášok. Počas semestra ich má 12, t.j. má 12 nezávislých pokusov doraziť na prednášku včas. Aká je pravdepodobnosť toho, že študent nestihne aspoň polovicu ranných prednášok v dôsledku toho, že zaspí?

Riešenie. Ak náhodná premenná X označuje počet prednášok, na ktoré študent zaspí, potom $P(X \geq 6) = P(X = 6 \vee X = 7 \vee X = 8 \vee X = 9 \vee X = 10 \vee X = 11 \vee X = 12) = 1 - P(X < 6) = 1 - P(X = 0 \vee X = 1 \vee X = 2 \vee X = 3 \vee X = 4 \vee X = 5) = 1 - F(5) = 1 - \text{BINOM.DIST}(5; 12; 0,3; 1) = 0,117849$.

Rozdelenie pravdepodobnosti náhodnej premennej X popisujúcej počet prednášok, na ktoré študent zaspí, je zobrazené na nasledujúcom obrázku 4.9.



Obr. 4.9 Pravdepodobnosť toho, že študent nestihne aspoň polovicu ranných prednášok v dôsledku toho, že zaspí

Úlohy

4.2 Študent má písať test, na ktorý sa nepripravil. Test obsahuje 25 otázok, odpovede majú formu áno – nie. Aby bol test úspešne absolvovaný, je potrebné označiť minimálne 18 správnych odpovedí. Aká je pravdepodobnosť, že študent bude úspešný, ak odpovede vyberá náhodne?

$$[=1-\text{BINOM.DIST}(17; 25; 0,5; 1)=0,021642625]$$

4.3 Test sa skladá z 12 otázok, za každou sú ponúknuté 4 odpovede, z ktorých je iba jedna správna. Aby študent urobil skúšku, musí správne odpovedať aspoň na 56% otázok testu. Aká je pravdepodobnosť, že študent urobí skúšku, ak:

- a) odpovede označuje úplne náhodne;
- b) v každej otázke vie vylúčiť jednu nesprávnu odpoveď, ale zo zvyšných si musí tipovať;
- c) je pripravený tak, že na 4 otázky pozná správnu odpoveď, ale odpovede na zvyšné otázky musí tipovať?

[a) $=1-\text{BINOM.DIST}(6;12;1/4;1)=0,014252782$; b) $=1-\text{BINOM.DIST}(6;12;1/3;1)=0,066448$; c) $=1-\text{BINOM.DIST}(2;8;1/4;1)=0,321457$]

4.4 Predpokladajme, že pravdepodobnosť narodenia dievčaťa je 0,49. Aká je pravdepodobnosť, že v rodine s tromi deťmi:

- a) nebude ani jedno dievča,
- b) bude viac chlapcov ako dievčat,
- c) budú všetky tri dievčatá?

[a) $=\text{BINOM.DIST}(0;3;0,49;0)=0,132651$; b) $=\text{BINOM.DIST}(0;3;0,49;0)+\text{BINOM.DIST}(1;3;0,49;0)=\text{BINOM.DIST}(1;3;0,49;1)=0,514998$; c) $=\text{BINOM.DIST}(3;3;0,49;0)=0,117649$]

4.5 Letecká spoločnosť prevádzkuje na určitej linke lietadlo pre 120 cestujúcich. I keď je na každý let všetkých 120 miest rezervovaných, priemerne 3% cestujúcich sa nedostaví a lietadlo často lietalo s prázdnyimi miestami. Spoločnosť preto rozhodla prijímať na konkrétny let rezervácie od 122 cestujúcich. Aká je pravdepodobnosť, že nebudú uspokojení všetci cestujúci, ktorí sa dostavia k letu?

[$=\text{BINOM.DIST}(121;122;0,97;0)+\text{BINOM.DIST}(122;122;0,97;0)=0,116134388$ alebo $=1-\text{BINOM.DIST}(120;122;1-0,03;1)=0,116134388$]

4.6 Predpokladáme, že sledované vzorky vody sú nezávislé vzhľadom k prítomnosti znečisťujúcej organickej látky. Ak vieme, že 8 zo 100 vzoriek obsahuje sledovanú nečistotu, aká bude pravdepodobnosť, že z vybraných 15 vzoriek budú obsahovať znečisťujúcu látku: a) presne 2, b) aspoň tri vzorky.

[a) $=\text{BINOM.DIST}(2;15;0,08;0)=0,227306$; b) $=1-\text{BINOM.DIST}(2;15;0,08;1)=0,112965$]

4.7 V elektrickom obvode je paralelne zapojených 24 žiaroviek. Je známe, že každá z nich je s pravdepodobnosťou 0,15 chybná. Určte pravdepodobnosť toho, že z týchto žiaroviek je:

- a) maximálne štvrtina chybných;
- b) minimálne jedna chybná;
- c) viac chybných, ako by sme mohli v priemere očakávať;
- d) menej chybných, ako by sme mohli s najväčšou pravdepodobnosťou očakávať.

[a) $=\text{BINOM.DIST}(6;24;0,15;1)=0,942808$; b) $=1-\text{BINOM.DIST}(0;24;0,15;0)=0,979767$; c) $E(X)=3,6=1-\text{BINOM.DIST}(3;24;0,15;1)=0,495121$; d) $=\text{BINOM.DIST}(2;24;0,15;1)=0,279828$]

4.8 Hodíme šesťkrát troma mincami. S akou pravdepodobnosťou nám padne jedna hlava s dvoma znakmi trikrát?

$$\left[= \text{BINOM.DIST}\left(3; 6; \frac{3}{8}; 0\right) = 0,257492 \right]$$

4.1.4 Poissonovo rozdelenie (Poisson distribution)

Toto diskkrétne rozdelenie vzniká ako limitný prípad binomického rozdelenia alebo vtedy, keď javy určitého druhu nastávajú náhodne v čase alebo v priestore. Je to rozdelenie počtu výskytov javu A za jednotku času, resp. špeciálne rozdelenie pre tzv. „maličké pravdepodobnosti“. Hovoríme, že náhodná premenná X má Poissonovo rozdelenie s parametrom $\lambda > 0$, ak nadobúda hodnoty $x = 0, 1, 2, \dots$ s pravdepodobnosťou

$$p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{pre } x = 0, 1, \dots \\ 0 & \text{inak} \end{cases}$$

Charakteristické vlastnosti tohto rozdelenia:

- pravdepodobnosť výskytu určitého javu v danom intervale (v čase alebo v priestore) je úmerná dĺžke tohto intervalu (napr. ak sa určitý jav vyskytne v priemere 12-krát za hodinu, tak za minútu sa vyskytne v priemere 0,2-krát);
- javy sa vyskytujú nezávisle v danom intervale a rovnako aj medzi po sebe nasledujúcimi intervalmi (sú výsledkom opakovaného systému nezávislých pokusov);
- priemer rozdelenia je rovný rozptylu, t.j. parametru λ .

$X \sim \text{Po}(\lambda)$	
$p(x)$	$= \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{pre } x = 0, 1, \dots, \\ 0 & \text{inak} \end{cases}$
Excel:	= POISSON.DIST(x; stred; kumulatívne) x je hodnota náhodnej premennej X ; stred je stredná hodnota λ ; kumulatívne je logická premenná – FALSE (0) určí, že počítame hodnotu pravdepodobnostnej funkcie, TRUE (1), ak počítame hodnotu distribučnej funkcie.

Číselné charakteristiky:

$E(X)$	$D(X)$	$\alpha(X)$	$\beta(X)$
λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$

Príklady Poissonovho rozdelenia pravdepodobnosti :

- počet ťažkých dopravných nehôd v určitom meste za deň;
- počet smrteľných pracovných úrazov v nejakom rezorte za rok;

- počet telefónnych hovorov v danom časovom intervale;
- počet častíc vyslaných rádioaktívnou látkou počas určitého časového intervalu;
- počet chýb na zvare pri nedeštruktívnej skúške;
- počet bublín na ploche danej veľkosti na odliatku;
- počet mikrodefektov na zadanej vzorke materiálu;
- počet trhlín na 1km kábla a pod.;
- výskyt porúch zariadenia v čase t ;
- počet zákazníkov v obchode za časovú jednotku, počet pracovných úrazov za rok, počet pacientov ošetrovaných v priebehu 1 dňa, počet červených krviniek v 1ml krvi;
- počet výskytov vzácneho ochorenia na konkrétnom území, počet narodených dvojčiat, trojčiat v určitej populácii...

Označuje sa ako zákon vzácných alebo zriedkavých javov (v čase alebo v priestore). Aj v binomickom rozdelení existujú zriedkavé javy, ale vždy sa uvažuje k nastaniu javu jeho nenastanie. Na niektoré javy sa však nedá aplikovať binomické rozdelenie: Napríklad, ak počas búrky bolo 5 bleskov, koľko bleskov nebolo? Ak cez križovatku prešlo od 9:00 do 10:00 deväťdesiat áut, koľko áut neprešlo? Ak sa v daný deň narodilo v danej nemocnici 5 detí, koľko sa nenarodilo?...

Príklad 4.6 Pri prevádzke výrobnjej linky vznikajú počas zmeny náhodné poruchy. Zo skúsenosti vieme, že v priebehu jednej zmeny dôjde v priemere k 2 poruchám. Aká je pravdepodobnosť, že v priebehu 24 hodín (trojzmennej prevádzky) nedôjde ani k jednej poruche?

Riešenie. Náhodná premenná X označuje počet porúch na výrobnjej linke počas 24 hodín. Predpokladáme, že X sa riadi Poissonovým rozdelením pravdepodobnosti, ktorého parameter λ je priamo úmerný dĺžke časového intervalu. Parameter λ predstavuje strednú hodnotu náhodnej premennej, t.j. stredný počet porúch počas 3 pracovných osemhodinových zmien, preto parameter $\lambda = 3 \cdot 2 = 6$ porúch.

$$P(X = 0) = \frac{6^0 e^{-6}}{0!} = \text{POISSON.DIST}(0; 6; 0) = 0,002478752$$

Pravdepodobnosť toho, že v priebehu 24 hodín nedôjde ani k jednej poruche je iba 2,5%.

Príklad 4.7 Predpokladajme, že v určitej populácii krýs sa albín vyskytuje s pravdepodobnosťou 0,001. Aká je pravdepodobnosť, že vo vzorke 100 krýs náhodne vybraných z tejto populácie:

- nebude ani jeden albín;
- bude práve jeden albín;
- bude najviac jeden albín;
- bude aspoň jeden albín;
- bude viac ako jeden albín.

Riešenie. Ak je pravdepodobnosť výskytu albína v danej populácii $p = 0,001$, potom predpokladaný počet albínov vo výberovej vzorke bude úmerný ich počtu $n = 100$, t.j. $\lambda =$

$100 \cdot 0,001 = 0,1$ (je to stredná hodnota náhodnej premennej $E(X) = np$). Ak náhodná premenná X označuje počet albínov, potom:

- a) $P(X = 0) = p(0) = \frac{0,1^0 e^{-0,1}}{0!} = 0,904837$
= POISSON.DIST(0;0,1;0);
- b) $P(X = 1) = p(1) = \frac{0,1^1 e^{-0,1}}{1!} = 0,090484$
= POISSON.DIST(1;0,1;0);
- c) $P(X = 0 \vee X = 1) = P(X = 0) + P(X = 1) = p(0) + p(1) = 0,995321$
= POISSON.DIST(1;0,1;1);
- d) $P(X > 0) = 1 - P(X = 0) = 1 - p(0) = 0,095163$
=1- POISSON.DIST(0;0,1;0);
- e) $P(X > 1) = 1 - (P(X = 0) + P(X = 1)) = 1 - p(0) - p(1) = 0,004679$
=1- POISSON.DIST(1;0,1;1).

Úlohy

4.9 Webovú stránku navštívi za sledované obdobie počas jednej hodiny v priemere 30 záujemcov (predpokladáme, že ich návštevy sú nezávislé). Určte pravdepodobnosť toho, že v priebehu 4 minút navštívi túto stránku:

- a) jeden návštevník,
- b) aspoň jeden návštevník,
- c) nie menej než traja ale menej ako desiatí návštevníci.

[a) = POISSON.DIST(1; 4 * 30/60; 0) = 0,270671; b) = 1 - POISSON.DIST(0; 2; 0) = 0,864665; c) = POISSON.DIST(9; 2; 1) - POISSON.DIST(2; 2; 1) = 0,323277]

4.10 K automatu s nápojmi príde v priemere 30 zákazníkov za 1/2 hodiny. Aká je pravdepodobnosť, že v priebehu nasledujúcich 5 minút:

- a) pristúpi k automatu práve 1 zákazník,
- b) prídu k automatu najviac 2 zákazníci,
- c) prídu k automatu aspoň 3 zákazníci?

[a) = POISSON.DIST(1; 5; 0) = 0,03369; b) = POISSON.DIST(2; 5; 1) = 0,124652; c) = 1 - POISSON.DIST(2; 5; 1) = 0,87348]

4.11 Cestovnú kanceláriu navštívi v priemere 5 osôb za hodinu. Aká je pravdepodobnosť, že v priebehu 15 minút nepríde do kancelárie nikto? Predpokladáme, že počet osôb, ktoré navštívia kanceláriu sa riadi Poissonovým rozdelením.

[= POISSON.DIST $\left(0; \frac{5}{4}; 0\right)$ = 0,286505]

4.12 Na infolinku príde v priebehu 8 hodín v priemere 650 žiadostí o spojenie. Aká je pravdepodobnosť, že v priebehu nasledujúcich 10 minút príde:

- a) 6 žiadostí o spojenie,
- b) najviac 6 žiadostí o spojenie?

[a) = POISSON.DIST $\left(6; \frac{650}{48}; 0\right)$ = 0,011262; b) = POISSON.DIST $\left(6; \frac{650}{48}; 1\right)$ = 0,018779]

4.13 V priebehu jednej hodiny príde na benzínové čerpadlo priemerne 90 zákazníkov. Určte pravdepodobnosť toho, že za 4 minúty prídu:

- a) práve dvaja zákazníci;
- b) najviac dvaja zákazníci;
- c) aspoň dvaja zákazníci;
- d) aspoň jeden zákazník.

[a) = POISSON.DIST(2; 6; 0) = 0,044618; b) = POISSON.DIST(2; 6; 1) = 0,061969; c) = 1 – POISSON.DIST(1; 6; 1) = 0,982649; d) = 1 – POISSON.DIST(0; 6; 0) = 0,997521]

4.14 K pokladnici prichádzajú priemerne dvaja návštevníci za minútu. Aká je pravdepodobnosť toho, že k pokladnici:

- a) v priebehu pol minúty príde aspoň 1 návštevník,
- b) v priebehu 10 minút nepríde nikto,
- c) počas hodiny príde aspoň 100 návštevníkov?

[a) = 1-POISSON.DIST(0;1;0)= 0,63212; b)=POISSON.DIST(0;20;0)= 2,06115E-09 ;c)

=1-POISSON.DIST(99;120;1)= 0,97213626]

4.15 Skúmaním bolo zistené, že na súbore 350 odlietkov je 406 bublín. Vyjadrite rozdelenie pravdepodobnosti počtu bublín na 1 odliatku.

$$[X \sim \text{Po} \left(\lambda = \frac{406}{350} = 1,16 \right), p(x) = \begin{cases} \frac{1,16^x e^{-1,16}}{x!} & \text{pre } x = 0, 1, \dots, 406 \\ 0 & \text{inak} \end{cases}]$$

4.16 Prijímač má za 10 000 hodín prevádzky v priemere 10 porúch. Určte pravdepodobnosť poruchy za 300 hodín prevádzky.

$[X \sim \text{Po} \left(\frac{10}{10000} \cdot 300 = 0,3 \right), = 1 - \text{POISSON.DIST}(0; 0,3; 0) = 0,259182]$

4.17 Kazovosť prevádzky pri výrobe televízorov je 8 kusov na 1000 výrobkov. Aká je pravdepodobnosť, že vo výrobnnej sérii 300 televízorov sa vyskytnú najviac 3 poruchové?

[= POISSON.DIST(3; 300 · 0,008; 1) = 0,77872]

4.1.5 Hypergeometrické rozdelenie (Hypergeometric distribution)

Z konečnej množiny N objektov (prvkov), medzi ktorými je M ($1 \leq M \leq N$) takých, ktoré majú určitú sledovanú vlastnosť, vyberieme náhodne n ($1 \leq n < N$) objektov. Objekty vyberáme naraz alebo postupne bez vrátenia, t.j. bez opakovania (závislé výbery – pravdepodobnosť úspechu v danom pokuse závisí od výsledku predchádzajúceho pokusu). Náhodná premenná X , ktorá predstavuje počet x jednotiek z n vybraných jednotiek, ktoré majú sledovanú vlastnosť, má hypergeometrické rozdelenie, označuje sa $X \sim Hg(N, M, n)$.

Zákon (funkcia) rozdelenia pravdepodobnosti:

$X \sim Hg(N, M, n)$	
$p(x) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} & \text{pre } \max\{0, n-N+M\} \leq x \leq \min\{n, M\}, x \in N_0, \\ 0 & \text{inak} \end{cases}$ $N, M, n \in N: 1 \leq n < N, 1 \leq M < N$	
Excel:	=HYPGEOM.DIST(vzorka_s; veľkosť_vzorky; populácia_s; počet_pop; kumulatívne) vzorka_s je hodnota náhodnej premennej X ; veľkosť_vzorky určuje veľkosť výberu, t.j. n ; populácia_s udáva počet prvkov so sledovanou vlastnosťou – parameter M ; počet_pop je celkový počet prvkov – parameter N ; kumulatívne priradíme logickú hodnotu 0, ak počítame hodnotu $p(x)$ a logickú hodnotu 1, ak chceme vypočítať hodnotu distribučnej funkcie.

Poznámka 4.1 Pri výbere nám nezáleží na poradí prvkov, preto $\binom{M}{x} \binom{N-M}{n-x}$ je počet priaznivých možností, t.j. počet možností, ako vybrať x prvkov s danou vlastnosťou a zároveň $(n-x)$ prvkov, ktoré uvedenú vlastnosť nemajú. $\binom{N}{n}$ počet možností výberu všetkých n prvkov.

Číselné charakteristiky:

$E(X)$	$D(X)$	$\alpha(X)$
$n \frac{M}{N}$	$n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$	$\frac{(N-2M)(N-1)^{\frac{1}{2}}(N-2n)}{nM(N-M)(N-n)^{\frac{1}{2}}(N-2)}$

Príklady hypergeometrického rozdelenia:

S hypergeometrickým rozdelením sa stretávame v štatistickej kontrole akosti výrobkov. Náhodná premenná X má hypergeometrické rozdelenie, ak jej hodnoty predstavujú napríklad počet chybných výrobkov medzi n náhodne vybranými výrobkami z dodávky.

Poznámka 4.2 Z porovnania charakteristík binomického a hypergeometrického rozdelenia môžeme usúdiť, že nastáva ich zhoda v prípade, že počet prvkov N v súbore je oproti počtu vybraných prvkov n veľmi veľký. Potom podiel $\frac{N-n}{N-1}$ sa približne rovná 1.

Teda ak je N vzhľadom k n dostatočne veľké (tzv. výberový podiel $\frac{n}{N} \leq 0,05$), možno hypergeometrické rozdelenie aproximovať binomickým rozdelením s parametrami n a $p = \frac{M}{N}$.

$$\left(\frac{n}{N} \leq 0,05\right) \Rightarrow Hg(N, M, n) \sim Bi\left(n, p = \frac{M}{N}\right)$$

Príklad 4.8 Určitý typ súčiastok je dodávaný v sérii po 500 kusoch. Pri preberacej kontrole je z každej série náhodne vybraných 5 výrobkov, ktoré sa skúškou znehodnotia. Séria je prijatá, ak medzi kontrolovanými výrobkami nie je žiadny chybný. Predpokladajme, že v sérii je 25 výrobkov chybných.

- Popíšte náhodnú premennú udávajúcu počet chybných vybraných výrobkov pomocou pravdepodobnostnej a distribučnej funkcie.
- S akou pravdepodobnosťou bude séria prijatá?
- Určte strednú hodnotu, rozptyl, smerodajnú odchýlku a modus počtu nepodarkov vo výbere.

Riešenie.



Náhodnú premennú X , ktorá označuje počet chybných výrobkov medzi vybranými piatimi výrobkami, môžeme popísať hypergeometrickým rozdelením $X \sim Hg(500, 25, 5)$. Takto definovaná náhodná premenná môže nadobúdať hodnoty z oboru hodnôt $\{0, 1, 2, 3, 4, 5\}$. Jej

pravdepodobnostná funkcia $p(x) = \begin{cases} \frac{\binom{25}{x} \binom{475}{5-x}}{\binom{500}{5}} & \text{pre } x = 0, 1, \dots, 5 \\ 0 & \text{inak} \end{cases}$.

Na výpočet jednotlivých kombinačných čísiel $\binom{n}{k}$ môžeme použiť v Exceli preddefinovanú funkciu `=COMBIN(počet;vybratý počet)`, kde argumentom „počet“ zadávame veľkosť čísla n a argumentom „vybratý počet“ veľkosť čísla k , obr. 4.10.

Hodnoty pravdepodobnostnej funkcie $p(x)$ je však možno vypočítať aj priamo pomocou funkcie `=HYPGEOM.DIST(vzorka_s; veľkosť_vzorky; populácia_s; počet_pop; kumulatívne)`, obr. 4.11.

Zákon rozdelenia pravdepodobnosti na nasledujúcom obr. 4.10.

	A	B	C	D	E	F	G
1	xi	0	1	2	3	4	5
2	pi	=COMBIN(25;0)* COMBIN(475;5)/ COMBIN(500;5)	=COMBIN(25;1)* COMBIN(475;4)/ COMBIN(500;5)	=COMBIN(25;2)* COMBIN(475;3)/ COMBIN(500;5)	=COMBIN(25;3)* COMBIN(475;2)/ COMBIN(500;5)	=COMBIN(25;4)* COMBIN(475;1)/ COMBIN(500;5)	=COMBIN(25;5)* COMBIN(475;0)/ COMBIN(500;5)

	A	B	C	D	E	F	G	H
1	xi	0	1	2	3	4	5	
2	pi	0,772962	0,205138	0,020862	0,001014	2,35E-05	2,08E-07	1

Obr. 4.10 Výpočet hodnôt pravdepodobnostnej funkcie $p(x)$ realizovaný v Exceli pomocou funkcie `COMBIN`

	A	B		A	B
1	xi	pi		1	xi
2	0	=HYPGEOM.DIST(A2;5;25;500;0)		2	0
3	1	=HYPGEOM.DIST(A3;5;25;500;0)		3	1
4	2	=HYPGEOM.DIST(A4;5;25;500;0)		4	2
5	3	=HYPGEOM.DIST(A5;5;25;500;0)		5	3
6	4	=HYPGEOM.DIST(A6;5;25;500;0)		6	4
7	5	=HYPGEOM.DIST(A7;5;25;500;0)		7	5
8	Σ	=SUM(B2:B7)		8	Σ

Obr. 4.11 Výpočet hodnôt pravdepodobnostnej funkcie $p(x)$ realizovaný v Exceli HYPGEOM.DIST

Hodnoty „skokov“ distribučnej funkcie $F(x)$ vypočítame pomocou kumulatívnej pravdepodobnosti (obr. 4.12) alebo pomocou funkcie =HYPGEOM.DIST s posledným argumentom rovným TRUE, t.j. 1 (obr. 4.13).

	A	B	C	D	E	F	G	H
1	xi	0	1	2	3	4	5	Σ
2	pi	0,772962	0,205138	0,020862	0,001014	2,35411E-05	2,08E-07	1
3	Fi	=B2	=B3+C2	=C3+D2	=D3+E2	=E3+F2	=F3+G2	

	A	B	C	D	E	F	G	H
1	xi	0	1	2	3	4	5	Σ
2	pi	0,772962	0,205138	0,020862	0,001014	2,35411E-05	2,08E-07	1
3	Fi	0,7729618	0,9781003	0,9989618	0,9999763	0,9999998	1,0000000	

Obr. 4.12 Výpočet hodnôt distribučnej funkcie $F(x)$ v Exceli vypočítaný kumulovaním jednotlivých pravdepodobností

	A	B	C	D	E
1	xi	pi	Fi		
2	0	0,772962	=HYPGEOM.DIST(A2;5;25;500;1)		
3	1	0,205138	=HYPGEOM.DIST(A3;5;25;500;1)		
4	2	0,020862	=HYPGEOM.DIST(A4;5;25;500;1)		
5	3	0,001014	=HYPGEOM.DIST(A5;5;25;500;1)		
6	4	2,35E-05	=HYPGEOM.DIST(A6;5;25;500;1)		
7	5	2,08E-07	=HYPGEOM.DIST(A7;5;25;500;1)		
8	Σ	1			

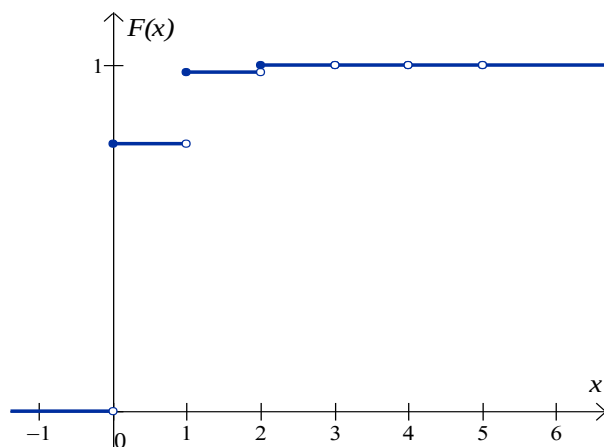
	A	B	C
1	xi	pi	Fi
2	0	0,772962	0,7729618
3	1	0,205138	0,9781003
4	2	0,020862	0,9989618
5	3	0,001014	0,9999763
6	4	2,35E-05	0,9999998
7	5	2,08E-07	1,0000000
8	Σ	1	

Obr. 4.13 Výpočet hodnôt distribučnej funkcie $F(x)$ v Exceli pomocou funkcie HYPGEOM.DIST

Graf distribučnej funkcie $F(x)$
s predpisom

$$F(x) = \begin{cases} 0 & \text{pre } x \in (-\infty, 0), \\ 0,7729618 & \text{pre } x \in \langle 0, 1), \\ 0,9781003 & \text{pre } x \in \langle 1, 2), \\ 0,9989618 & \text{pre } x \in \langle 2, 3), \\ 0,9999763 & \text{pre } x \in \langle 3, 4), \\ 0,9999998 & \text{pre } x \in \langle 4, 5), \\ 1 & \text{pre } x \in \langle 5, \infty) \end{cases}$$

je na obr. 4.14.



Obr. 4.14 Graf $F(x)$ pre $X \sim Hg(500, 25, 5)$

Pravdepodobnosť toho, že séria bude prijatá, sa rovná pravdepodobnosti toho, že medzi kontrolovanými výrobkami nebude žiadny chybný výrobok, t.j. počítame $p(X = 0) = 0,772962$.

Výpočet strednej hodnoty, rozptylu, smerodajnej odchýlky možno urobiť priamo v Exceli, obr. 4.15. Modus počtu nepodarkov vo výbere je rovný 0, je to hodnota náhodnej premennej, ktorá sa vyskytuje s najväčšou pravdepodobnosťou.

	A	B	C	D	E		A	B	C	D	E
1	xi	pi	Fi	xi*pi	(xi-E(X))^2*pi		xi	pi	Fi	xi*pi	(xi-E(X))^2*pi
2	0	0,772962	0,7729618	=A2*B2	=(A2-\$D\$8)^2*B2		0	0,772962	0,7729618	0	0,048310113
3	1	0,205138	0,9781003	=A3*B3	=(A3-\$D\$8)^2*B3		1	0,205138	0,9781003	0,205138	0,115390398
4	2	0,020862	0,9989618	=A4*B4	=(A4-\$D\$8)^2*B4		2	0,020862	0,9989618	0,041723	0,063888469
5	3	0,001014	0,9999763	=A5*B5	=(A5-\$D\$8)^2*B5		3	0,001014	0,9999763	0,003043	0,007671468
6	4	2,35E-05	0,9999998	=A6*B6	=(A6-\$D\$8)^2*B6		4	2,35E-05	0,9999998	9,42E-05	0,000331047
7	5	2,08E-07	1,0000000	=A7*B7	=(A7-\$D\$8)^2*B7		5	2,08E-07	1,0000000	1,04E-06	4,69646E-06
8	Σ	1		=SUM(D2:D7)	=SUM(E2:E7)		Σ	1		0,25	0,235596192
9											
10	E(X)=	=D8	=5*25/500				E(X)=	0,25	0,25		
11	D(X)=	=E8	=5*25/500*(1-25/500)*((500-5)/(500-1))				D(X)=	0,235596	0,235596		
12	σ(X)=	=SQRT(B11)					σ(X)=	0,485383			
13	Mo(X)=	0					Mo(X)=	0			

Obr. 4.15 Výpočet $E(X)$, $D(X)$, $\sigma(X)$ realizovaný v Exceli

Úlohy

4.18 Medzi 4 000 závitov M10 bolo omylom primiešaných 1 000 závitov M9. Aké bude rozdelenie pravdepodobnosti, že pri náhodnom výbere 5 závitov bude 0,1,2,3,4,5 správneho rozmeru? Ak pri montáži výrobku potrebuje pracovník 4 závity s rozmerom M10, aká bude pravdepodobnosť, že medzi vybranými 5 závitmi budú aspoň 4 s požadovanými vlastnosťami? Koľko závitov správneho rozmeru vytiahne s najväčšou pravdepodobnosťou?

$$\left[p(x) = \begin{cases} \binom{4000}{x} \binom{1000}{5-x} / \binom{5000}{5} & \text{pre } x = 0, 1, \dots, 5; \\ 0 & \text{inak} \end{cases}; P(X \geq 4) = 1 - \right. \\ \left. \text{HYPGEOM.DIST}(3; 5; 4000; 5000; 1) = 0,737321; 2 \right]$$

4.19 Štatisticky bolo zistené, že v závode sa vyrobí v priemere na každých 100 výrobkov 5 nepodarkov. Nech náhodná premenná udáva počet nepodarkov v dodávke 10 kusov. Aká je pravdepodobnosť, že v dodávke:

- bude najviac 1 nepodarok,
- bude viac než 1 nepodarok,
- bude práve 1 nepodarok?
- Určte strednú hodnotu, modus, rozptyl, smerodajnú odchýlku sledovanej náhodnej premennej.

$$[X \sim Hg(100, 5, 10); a) = \text{HYPGEOM.DIST}(1; 10; 5; 100; 1) = 0,923; b) = 1 - \text{HYPGEOM.DIST}(1; 10; 5; 100; 1) = 0,077; c) = \text{HYPGEOM.DIST}(1; 10; 5; 100; 0) = 0,339; d) 0,5; 0; 0,432; 0,657]$$

4.20 V dodávke 85 obrobkov sú 4 kusy chybné. Náhodne vyberieme naraz 5 obrobkov. Aká je pravdepodobnosť, že medzi vybranými obrobkami budú najviac 2 chybné?

$$\left[X \sim Hg(85, 4, 5); P(X \leq 2) = \frac{\text{COMBIN}(4; 0) * \text{COMBIN}(81; 5)}{\text{COMBIN}(85; 5)} + \frac{(\text{COMBIN}(4; 1) * \text{COMBIN}(81; 4))}{\text{COMBIN}(85; 5)} + \right. \\ \left. \frac{\text{COMBIN}(4; 2) * \text{COMBIN}(81; 3)}{\text{COMBIN}(85; 5)} = 0,999602 \text{ alebo } \text{HYPGEOM.DIST}(2; 5; 4; 85; 1) = 0,999602 \right]$$

4.21 Zo sady 50 výrobkov, medzi ktorými je 11 chybných, náhodne vyberieme na kontrolu kvality osem výrobkov. Určte pravdepodobnosť toho, že medzi kontrolovanými je:

- päť chybných výrobkov,
- aspoň jeden chybný výrobok,
- menej než päť chybných výrobkov.

Aký je najpravdepodobnejší a očakávaný (priemerný) počet vybraných chybných výrobkov?

$$[a) = \text{HYPGEOM.DIST}(5; 8; 11; 50; 0) = 0,007864; b) = 1 - \text{HYPGEOM.DIST}(0; 8; 11; 50; 0) = 0,885405; c) = \text{HYPGEOM.DIST}(4; 8; 11; 50; 1) = 0,991474; Mo = 2; E(X) = 1,76]$$

4.2 NIEKTORÉ ROZDELENIA SPOJITEJ NÁHODNEJ PREMENEJ

4.2.1 Rovnomerné (spojité) rozdelenie (Uniform distribution (continuous))

Rovnomerné rozdelenie $R(\alpha, \beta)$ sa používa ako model rozdelenia pravdepodobnosti náhodnej premennej vtedy, keď má náhodná premenná konštantnú hustotu pravdepodobnosti na intervale (α, β) , kde $\alpha, \beta \in R$.

Hustota a distribučná funkcia rovnomerného rozdelenia má tvar:

$X \sim R(\alpha, \beta)$	
$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{pre } x \in (\alpha, \beta) \\ 0 & \text{inak} \end{cases}$	$F(x) = \begin{cases} 0 & \text{pre } x \in (-\infty, \alpha) \\ \frac{x - \alpha}{\beta - \alpha} & \text{pre } x \in (\alpha, \beta) \\ 1 & \text{pre } x \in (\beta, \infty) \end{cases}$

Číselné charakteristiky:

$E(X)$	$D(X)$	$\alpha(X)$	$\beta(X)$
$\frac{\alpha + \beta}{2}$	$\frac{(\beta - \alpha)^2}{12}$	0	-1, 2

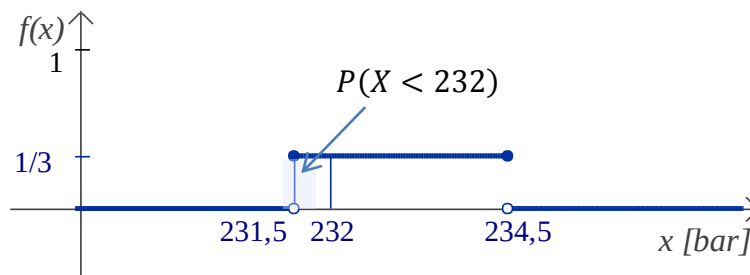
Týmto rozdelením sa riadi napríklad:

- doba čakania na uskutočnenie javu, ktorý sa opakuje v pravidelných intervaloch, (napr. doba čakania na dodávku tovaru, ktorá sa pravidelne opakuje alebo doba čakania na spoj, ktorý premáva v pravidelných intervaloch);
- veľkosť chyby pri zaokrúhľovaní čísel;
- chyby pri odčítaní údajov z meracích prístrojov (s lineárnou stupnicou).

Príklad 4.9 Výrobca má podľa normy plniť potápačské fľaše na stlačený vzduch plniacim tlakom 232 barov. V praxi je použitý skutočný plniaci tlak v rozsahu od 231,5 do 234,5 barov. Aká je pravdepodobnosť, že potápačská fľaša bude naplnená na menej ako požadovaných 232 barov, ak predpokladáte, že proces plnenia má rovnomerné rozdelenie?

Riešenie.

Ak predpokladáme $X \sim R(231,5; 234,5)$, potom funkcia hustoty pravdepodobnosti $f(x) = \begin{cases} \frac{1}{234,5 - 231,5} = 0, \bar{3} & \text{pre } x \in (231,5; 234,5) \\ 0 & \text{inak} \end{cases}$, t.j. pravdepodobnosť, že náhodná premenná X popisujúca plniaci tlak fľaše bude menšia ako 232 barov, je rovná plošnému obsahu prislúchajúceho obdĺžnika $P(X < 232) = 0, \bar{3} \cdot (232 - 231,5) = 0,1 \bar{6}$ (obr. 4.16), resp. vyjadrená pomocou distribučnej funkcie $F(232) = \frac{232 - 231,5}{234,5 - 231,5} = 0,1 \bar{6}$.



Obr. 4.16 Funkcia hustoty pravdepodobnosti rozdelenia $R(231,5; 234,5)$

Úlohy

- 4.22** Výrobná linka má plniť plechovky s nápojom 12 jednotkami sýtiaceho plynu. V skutočnosti ich ale plní v rozsahu od 11,5 do 12,5 jednotiek, pričom predpokladáte, že tento proces má rovnomerné rozdelenie. Aká je pravdepodobnosť, že v plechovke je menej ako 11,8 jednotiek plynu ?

[0,3]

- 4.23** Náhodná premenná X má distribučnú funkciu $F(x) = \begin{cases} 0 & \text{pre } x < -1 \\ \frac{x+1}{2} & \text{pre } -1 \leq x \leq 1 \\ 1 & \text{pre } x > 1 \end{cases}$.

Určte funkciu hustoty pravdepodobnosti, strednú hodnotu, rozptyl, medián, tretí decil a šesťdesiaty percentil tejto náhodnej premennej. Nájdite pravdepodobnosť, že náhodná premenná bude nadobúdať hodnoty väčšie ako 0,2; menšie ako 0,7; od -2 do -0,5; od -0,2 do 0,5.

$$\left[f(x) = \begin{cases} \frac{1}{2} & \text{pre } x \in (-1,1) \\ 0 & \text{inak} \end{cases} ; 0; \frac{1}{3}; 0; -0,6; 0,2; 0,4; 0,85; 0,25; 0,35 \right]$$

- 4.24** Autobusy MHD jazdia v pravidelných intervaloch po 25 minútach. Cestujúci príde na zastávku v ľubovoľnom okamihu. Sledujeme náhodnú premennú predstavujúcu dobu čakania na príchod autobusu. Popíšte túto náhodnú premennú pomocou funkcie hustoty a distribučnej funkcie. Vypočítajte pravdepodobnosti toho, že cestujúci bude čakať na spoj najviac 5 minút, práve 15 minút, najmenej 10 minút, 5 až 15 minút. Nájdite strednú hodnotu, medián, rozptyl, smerodajnú odchýlku a deviaty decil doby čakania na autobus.

$$\left[f(x) = \begin{cases} \frac{1}{25} & \text{pre } x \in (0,25) \\ 0 & \text{inak} \end{cases} ; F(x) = \begin{cases} 0 & \text{pre } x \leq 0 \\ \frac{x}{25} & \text{pre } 0 < x < 25 \\ 1 & \text{pre } x \geq 25 \end{cases} ; 0,2; 0; 0,6; 0,4; \right]$$

12,5; 12,5; 52,083; 7,22; 22,5]

4.25 Firma očakáva komponenty od dodávateľa v dobe od 12 do 24 hodín. Uskutočnenie dodávky je rovnako možné kedykoľvek v priebehu tohto intervalu. Označme ako náhodnú premennú dobu čakania na dodávku.

- Popíšte túto náhodnú premennú pomocou funkcie hustoty a distribučnej funkcie, funkcie vyjadrite predpisom i graficky.
- Nájdite pravdepodobnosti, že predajňa bude čakať na dodávku najviac 40 minút; práve 1 hodinu; minimálne ½ hodinu; minimálne 20 minút; najviac ale 80 minút.

$$[a) f(x) = \begin{cases} \frac{1}{12} & \text{pre } x \in (0,12) \\ 0 & \text{inak} \end{cases}; F(x) = \begin{cases} 0 & \text{pre } x \leq 0 \\ \frac{x}{12} & \text{pre } 0 < x < 12 \\ 1 & \text{pre } x \geq 12 \end{cases} ; b) 0,333; 0; 0,75; 0,5]$$

4.26 Pri vážení nemáme závažie menšie ako 1 gram. Ak výsledok váženia ukáže hmotnosť medzi 15 a 16 gramov, odhadneme hmotnosť na 15,5 gramov. Tento odhad je zaťažený chybou, ktorá je náhodnou premennou s konštantnou hustotou pravdepodobnosti v intervale od -0,5 do 0,5 gramov.

- Určte hustotu pravdepodobnosti, distribučnú funkciu, strednú hodnotu a rozptyl chyby odhadu.
- Aká je pravdepodobnosť, že absolútna chyba tohto odhadu neprekročí hodnotu 0,3 g?

$$[a) f(x) = \begin{cases} 1 & \text{pre } -0,5 < x < 0,5 \\ 0 & \text{inak} \end{cases}; F(x) = \begin{cases} 0 & \text{pre } x \leq -0,5 \\ x + 0,5 & \text{pre } -0,5 < x < 0,5 \\ 1 & \text{pre } x \geq 0,5 \end{cases};$$

0; 0,083; b) 0,6]

4.27 Pri zaokrúhľovaní čísla na 1 desatinné miesto sa dopúšťame chyby, ktorá je náhodnou premennou s rovnomerným rozdelením.

- Určte parametre α a β tohto rozdelenia a popíšte náhodnú premennú pomocou funkcie hustoty.
- Aká je pravdepodobnosť, že absolútna chyba pri zaokrúhľovaní neprekročí hodnotu 0,03?

$$[a) -0,05; 0,05; f(x) = \begin{cases} \frac{1}{0,1} & \text{pre } -0,05 < x < 0,05 \\ 0 & \text{inak} \end{cases}; b) 0,6]$$

4.2.2 Normálne (Gaussovo) rozdelenie (Normal (Gaussian) distribution)

Patrí medzi najčastejšie a najdôležitejšie rozdelenie pravdepodobnosti, pretože mnohé náhodné premenné v technike, prírodných, ekonomických a spoločenských vedách majú toto rozdelenie alebo sa ním dajú aproximovať. Má mimoriadny význam v teórii pravdepodobnosti a v matematickej štatistike. Klasickým typom veličín, ktoré sa riadia týmto rozdelením sú náhodné chyby, preto sa toto rozdelenie niekedy označuje ako zákon chýb. V anglickej literatúre je zaužívané pomenovanie odvodené od zvonovitého tvaru krivky - bell curve. Normálne rozdelenie je závislé od dvoch parametrov μ a σ , označuje sa $N(\mu, \sigma^2)$, kde μ je stredná hodnota a σ^2 je rozptyl náhodnej premennej.

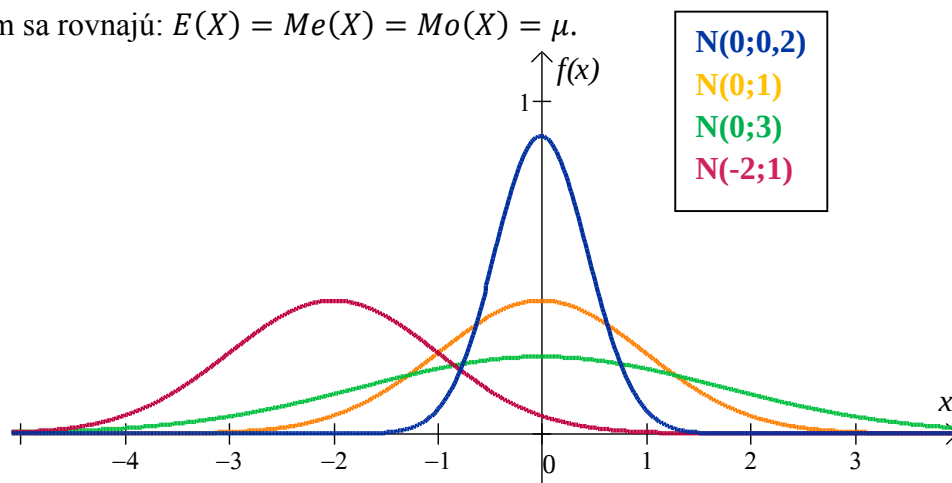
Hustota rozdelenia a distribučná funkcia majú tvar:

$X \sim N(\mu, \sigma^2)$	
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; x \in R$	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt; x \in R$
Excel:	=NORM.DIST(x;stred;smerodajná_odch;kumulatívne) x je hodnota náhodnej premennej X; stred je stredná hodnota (priemer) μ daného rozdelenia; smerodajná_odch je smerodajná odchýlka σ; môžeme voliť medzi frekvenčnou a distribučnou funkciou: kumulatívne zadávame 0 pre funkciu hustoty pravdepodobnosti $f(x)$ a 1 pre distribučnú funkciu $F(x)$.

Číselné charakteristiky:

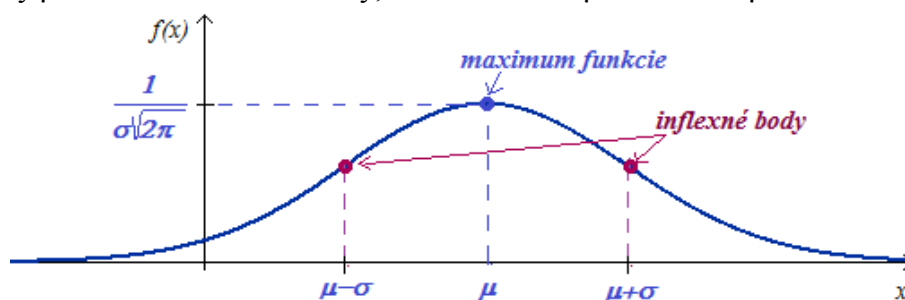
$E(X)$	$D(X)$	$\alpha(X)$	$\beta(X)$
μ	σ^2	0	0

Poznámka 4.3 Stredná hodnota, medián a modus náhodnej premennej s normálnym rozdelením sa rovnajú: $E(X) = Me(X) = Mo(X) = \mu$.

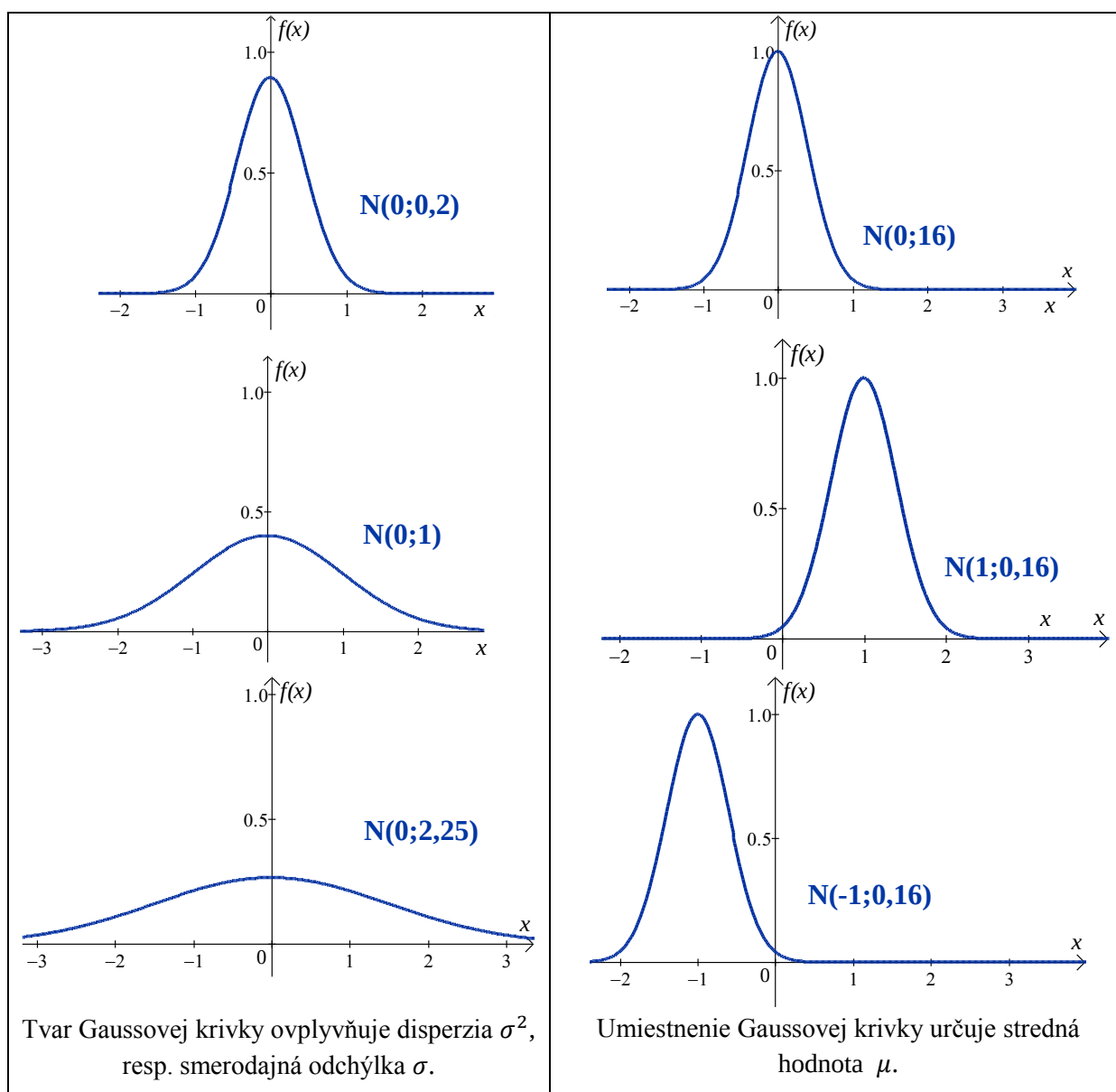


Obr. 4.17 Grafy funkcie hustoty pravdepodobnosti pre vybrané hodnoty parametrov μ a σ^2

Graf funkcie hustoty (obr. 4.18) má tvar zvona a je symetrický podľa osi $x = \mu$ (zo symetrie vyplýva $P(X > \mu) = P(X < \mu) = 0,5$), parameter μ teda určuje, kde má krivka maximum s hodnotou $\frac{1}{\sigma\sqrt{2\pi}}$, parameter σ^2 je jediným parametrom, ktorý určuje tvar krivky, t.j. určuje ako sú od hodnoty μ vzdialené inflexné body, ktoré sú v $x = \mu + \sigma$ a $x = \mu - \sigma$.



Obr. 4.18 Gaussova krivka - graf hustoty rozdelenia pravdepodobnosti normálne rozdelenej náhodnej premennej $X \sim N(\mu, \sigma^2)$



Obr. 4.19 Grafy funkcie hustoty pravdepodobnosti pre vybrané hodnoty parametrov μ a σ^2

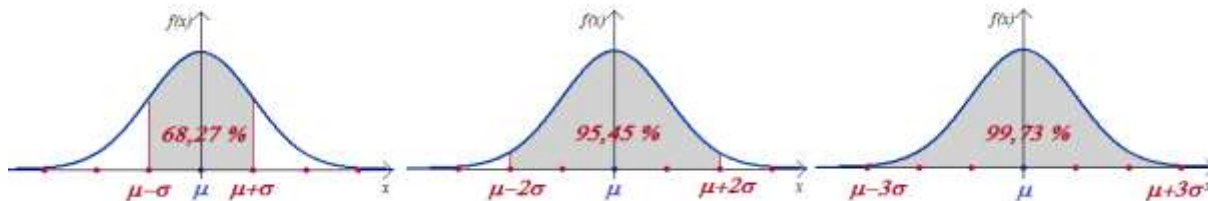
Veta 4.1 Pravidlo troch sigma

Pre náhodnú premennú $X \sim N(\mu, \sigma^2)$ platí:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0,6827$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0,9545$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0,9973$$



Obr. 4.20 Pravidlo troch sigma

Ak predpokladáme, že náhodná premenná X veľmi rozsiahleho súboru má normálne rozdelenie, pre ktoré sme zostrojili normálnu krivku, použitím pravidla troch sigma môžeme tvrdiť, že interval $(\mu - \sigma, \mu + \sigma)$ vymedzuje plochu veľkosti 0,6827 (t.j. viac ako 2/3 plochy), a teda táto premenná má 68,27% hodnôt v intervale $(\mu - \sigma, \mu + \sigma)$. Môžeme tiež tvrdiť, že ľubovoľný prvok nášho súboru padne s pravdepodobnosťou 68,27% do intervalu $(\mu - \sigma, \mu + \sigma)$ a s pravdepodobnosťou 31,73% mimo tohto intervalu. Analogicky môžeme konštatovať aj pre ďalšie dva intervaly $\mu \pm 2\sigma$, $\mu \pm 3\sigma$.

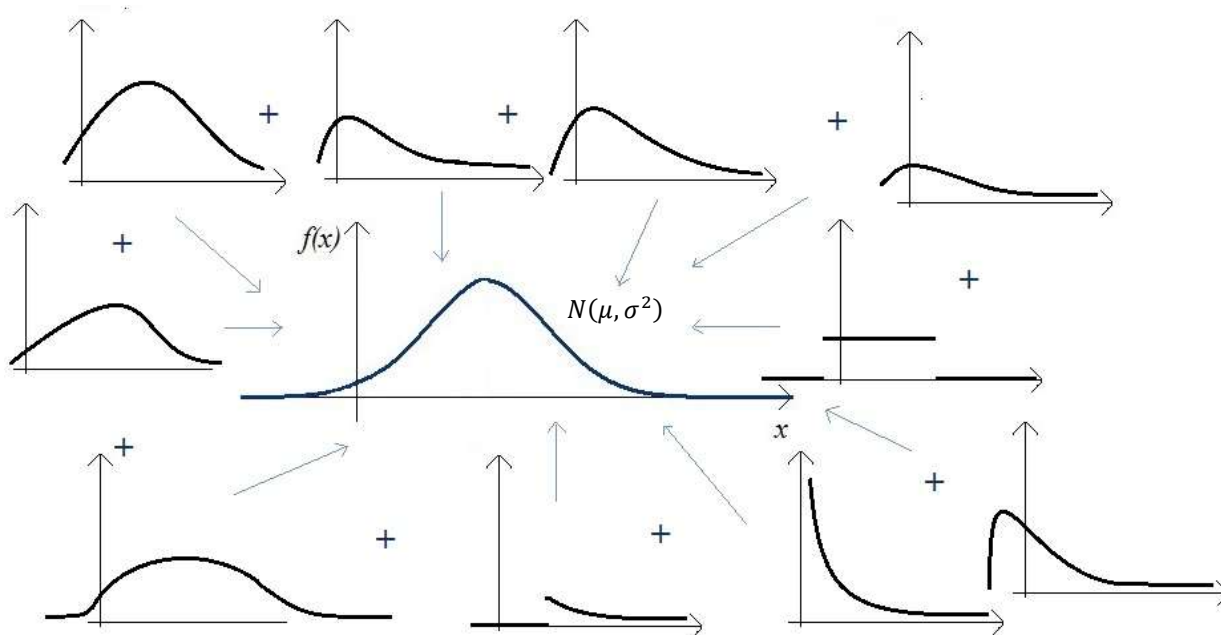
V praxi sa pravidlo troch sigma používa na prvý odhad štandardnej odchýlky pre náhodnú premennú, u ktorej predpokladáme normálne rozdelenie. Jej hodnotu dostaneme, ak rozpätie, t.j. rozdiel najväčšej a najmenšej hodnoty náhodnej premennej vydáme šiestimi.

Zaujímavosť:

Chvost normálneho rozdelenia je oblasť napravo od určitého bodu na pravej strane od strednej hodnoty alebo symetricky naľavo na ľavej strane krivky. Pomenovanie „chvost“ má pôvod v obrázku W. Gosseta z r.1908.



Toto rozdelenie vystihuje rozdelenie spojitých kvantitatívnych premenných. Je vo všeobecnosti vhodné k popisu všetkých náhodných premenných, ktoré je možné interpretovať ako aditívny výsledok mnohých nepatrných a vzájomne nezávislých vplyvov. Stačí, aby sčítancov bol dostatočný počet a aby žiadny z nich nemal na výslednú náhodnú premennú rozhodujúci vplyv. Platí, že čím viac nezávislých náhodných premenných sčítame, tým sa ich súčet viac blíži k normálnemu rozdeleniu bez ohľadu na to, aké mali pôvodné premenné rozdelenie, obr. 4.21.



Obr. 4.21 Normálne rozdelenie je zákonom rozdelenia súčtu ľubovoľných náhodných premenných

Podľa centrálnej limitnej vety náhodná premenná, ktorá je určená pôsobením väčšieho počtu navzájom nezávislých vplyvov, má aspoň približne normálne rozdelenie, bez ohľadu na to, ako je každý z týchto faktorov rozdelený.

Príklady použitia normálneho rozdelenia pravdepodobnosti:

- veľkosť chyby meraní;
- rozdelenie skutočného rozmeru vnútri tolerančného poľa pri obrábaní - odchýlka rozmeru výrobku od požadovanej hodnoty;
- spracovanie signálu, spracovanie obrazu;
- biometrické veličiny (v praxi sa stretávame aj s rôznymi modifikáciami Gaussovho rozdelenia, napr. s logaritmicko-normálnym rozdelením, ktoré sa vyznačuje kladným zošikmením – popisuje rozdelenie koncentrácie, hmotnosti ľudí, Ex-Gaussovým rozdelením popisujúcim rozdelenie reakčných časov...);
- v štatistike - aby bolo možné hypotézy pre ľubovoľnú náhodnú premennú, ktorá sa riadi týmto rozdelením verifikovať, používajú sa hodnoty normovaného normálneho rozdelenia, ktoré označujeme $N(0,1)$.

4.2.2.1 Štandardné (štandardizované, normované) normálne rozdelenie (Z-rozdelenie) (standard normal distribution or the unit normal distribution)

Ak náhodná premenná X má normálne rozdelenie s parametrami μ a σ , potom náhodnú premennú $Z = \frac{X-\mu}{\sigma}$ nazývame náhodná premenná s normovaným normálnym rozdelením. Platí $E(Z) = 0$ a $D(Z) = 1$. Toto rozdelenie sa označuje $N(0,1)$. Normované normálne rozdelenie sa používa ako základné, s ktorým sa porovnávajú iné rozdelenia. Funkcia hustoty

pravdepodobnosti sa v literatúre a tabuľkách označuje $\varphi(z)$, Excel používa značenie $f(z)$, jeho distribučná funkcia sa označuje $\Phi(z)$, resp. Excel $F(z)$.



Obr. 4.22 Pri štandardizácii $Z = \frac{x-\mu}{\sigma}$ odčítaním μ posúvame krivku do nuly, delením σ meníme tvar krivky, avšak plocha pod krivkou zostáva stále rovná 1

Funkčný predpis hustoty a distribučnej funkcie:

$Z \sim N(0, 1)$	
$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}; z \in R$	$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt; z \in R$
Excel:	=NORM.S.DIST(z;kumulatívne) z je hodnota normovanej náhodnej premennej; kumulatívne môže nadobúdať hodnotu 0 pre $\varphi(z)$ a 1 pre hodnotu $\Phi(z)$.

Číselné charakteristiky:

$E(Z)$	$D(Z)$	$\alpha(Z)$	$\beta(Z)$
0	1	0	0

Vzťahy medzi ľubovoľným $N(\mu, \sigma^2)$ a $N(0, 1)$:

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right);$$

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

$$P(a \leq X \leq b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Veta 4.2 Vlastnosti rozdelenia $N(0, 1)$:

- $\varphi(z)$ je párna funkcia, t.j. pre $\forall z \in R: \varphi(-z) = \varphi(z)$;
- v $z = 0$ má funkcia hustoty pravdepodobnosti maximum $\varphi(0) = \frac{1}{\sqrt{2\pi}} \doteq 0,4$;
- $\Phi(-z) = 1 - \Phi(z)$.

Poznámka 4.4 Vo väčšine štatistických analýz a testov je hlavným predpokladom normálne rozdelenie dát.

Kvantily normálneho rozdelenia

Kvantily normálneho rozdelenia budeme označovať x_α a normovaného normálneho rozdelenia z_α , kde α je pravdepodobnosť „naľavo“ od x_α , resp. z_α (veľkosť „zľava ohraničenej“ plochy pod hustotou).

V Exceli používame pre hľadanie kvantilov funkciu:

- pre ľubovoľné normálne rozdelenie $N(\mu, \sigma^2)$ a hľadanú pravdepodobnosť α =NORM.INV(pravdepodobnosť;stred;smerodajná_odch), kde argument pravdepodobnosť je pravdepodobnosť α , pre ktorú hľadáme kvantil x_α ; stred je stredná hodnota μ daného rozdelenia; smerodajná_odch je smerodajná odchýlka σ ;
- pre normované normálne rozdelenie $N(0, 1)$ a hľadanú pravdepodobnosť α =NORM.S.INV(pravdepodobnosť). Zadáваме iba pravdepodobnosť α , pre ktorú hľadáme kvantil z_α .

Príklad 4.10 Letecká spoločnosť sa snaží optimalizovať množstvo tankovaného paliva na pravidelnej linke. Dlhodobým pozorovaním bolo zistené, že spotreba paliva, v závislosti na letových podmienkach a obsadenosti lietadla, má normálne rozdelenie so strednou hodnotou 5,9 tony a smerodajnou odchýlkou 0,5 tony. Aké množstvo paliva je potrebné na to, aby lietadlo doletelo do cieľového miesta s pravdepodobnosťou 99% bez nutnosti medzipristátia z dôvodu doplnenia paliva?

Riešenie. $X \sim N(5,9; 0,5^2)$

$$P(X \leq x_\alpha) = F(x_\alpha) = 0,99 \Rightarrow x_\alpha = F^{-1}(0,99) = \text{NORM.INV}(0,99; 5,9; 0,5) = 7,063174$$

Lietadlo potrebuje natankovať 7,06 ton paliva.

Príklad 4.11 Každé meranie je sprevádzané náhodnými a systematickými chybami. Meranie ampérmetrom je zaťažené systematickou chybou 20 mA, náhodné chyby merania majú normálne rozdelenie pravdepodobnosti so smerodajnou odchýlkou 5mA. S akou pravdepodobnosťou sa bude líšiť chyba nameranej hodnoty pri jednom meraní maximálne o 12 mA od skutočnej meranej hodnoty? Určte hodnotu meranej veličiny, ktorá nebude s pravdepodobnosťou 0,9 prekročená.

Riešenie. X je náhodná premenná, ktorá nadobúda hodnoty chyby pri jednom meraní daným ampérmetrom. Systematická chyba je priemernou chybou: $E(X) = \mu = 20$ mA, smerodajná odchýlka $\sigma = 5$ mA. Platí $X \sim N(0 + 20, 5^2)$.

Pravdepodobnosť toho, že chyba nameranej hodnoty pri jednom meraní sa bude líšiť maximálne o 12mA, t.j. neprekročí v absolútnej hodnote 12 mA bude $P(|X| \leq 12) = P(-12 \leq X \leq 12) = \text{NORM.DIST}(12; 20; 5; 1) - \text{NORM.DIST}(-12; 20; 5; 1) = 0,054799$.

Hodnotu meranej veličiny, ktorá nebude s pravdepodobnosťou 0,9 prekročená určíme ako $P(|X| \leq a) = 0,9 \Rightarrow a = \text{NORM.INV}(0,9; 20; 5) = 26,40776$.

Úlohy

4.28 Nech X má normálne rozdelenie so strednou hodnotou 15 a smerodajnou odchýlkou 4. Určte nasledovné pravdepodobnosti :

- a) $P(X = 12)$; b) $P(X < 12)$; c) $P(X > 15)$;
d) $P(10 < X < 20)$; e) $P(|X| < 9)$; f) $P(|X| > 9)$.

[a] 0; b) = NORM.DIST(12; 15; 4; 1) = 0,226627 ; c) $1 - \text{NORM.DIST}(15; 15; 4; 1) = 0,5$; d) = NORM.DIST(20; 15; 4; 1) – NORM.DIST(10; 15; 4; 1) = 0,7887; e) = NORM.DIST(9; 15; 4; 1) – NORM.DIST(–9; 15; 4; 1) = 0,066807; f) = NORM.DIST(–9; 15; 4; 1) + 1 – NORM.DIST(9; 15; 4; 1) = 0,9331928]

4.29 Nech X má normálne rozdelenie so strednou hodnotou 5 a rozptylom 4. Určte nasledovné pravdepodobnosti :

- a) $P(X = 5)$; b) $P(X < 5)$; c) $P(X > 5)$;
d) $P(4 < X < 5)$; e) $P(|X| < 5)$; f) $P(|X| > 4)$.

[a] 0; b) 0,5 ; c) 0,5; d) = NORM.DIST(5; 5; 2; 1) – NORM.DIST(4; 5; 2; 1) = 0,191462 ; e) = NORM.DIST(5; 5; 2; 1) – NORM.DIST(–5; 5; 2; 1) = 0,5; f) = 1 – (NORM.DIST(4; 5; 2; 1) – NORM.DIST(–4; 5; 2; 1)) = 0,691465859]

4.30 Náhodná premenná X má normálne rozdelenie so strednou hodnotou 6 a rozptylom rovným 0,81. Určte pravdepodobnosť, že táto náhodná veličina nadobudne hodnotu:

- a) maximálne 5; b) aspoň 5,5; c) z intervalu (5,7).

[a] = NORM.DIST(5; 6; 0,9; 1) = 0,13326; b) $1 - \text{NORM.DIST}(5,5; 6; 0,9; 1) = 0,710743$; c) = (NORM.DIST(7; 6; 0,9; 1) – (NORM.DIST(5; 6; 0,9; 1) = 0,733479]

4.31 Nech Z má normované normálne rozdelenie. Nájdite hodnotu z , ktorá spĺňa nasledovné rovnice:

- a) $P(Z < z) = 0,95$; b) $P(Z < z) = 0,5$; c) $P(Z > z) = 0,05$;
d) $P(Z > z) = 0,9$; e) $P(-1,23 < Z < z) = 0,85$; f) $P(z < Z < 1,92) = 0,95$.

[a] = NORM.S.INV(0,95) = 1,644854; b) = NORM.S.INV(0,5) = 0 ; c) = NORM.S.INV(1 – 0,05) = 1,644854; d) = NORM.S.INV(0,1) = –1,28155; e) NORM.S.INV(NORM.S.DIST(–1,23; 1) + 0,85) = 1,743176; f) = NORM.S.INV(NORM.S.DIST(1,92; 1) – 0,95) = –2,00333]

4.32 Je daná náhodná premenná $X \sim N(2, 3^2)$. Aké veľké musí byť číslo $\check{c}$, aby hodnoty premennej X padli do intervalu ($\check{c}$, 3) s pravdepodobnosťou aspoň 0,25?

Určte medián $x_{0,5}$.

[maximálne 1,087958372 pozn.: $P(\check{c} \leq X \leq 3) = F(3) - F(\check{c}) \geq 0,25 \Rightarrow F(\check{c}) \leq F(3) - 0,25$; = NORM.DIST(3; 2; 3; 1) – 0,25 $\rightarrow F(\check{c}) = 0,38055866$; = NORM.INV(0,38055866; 2; 3) $\rightarrow \check{c} = 1,087958372$; $x_{0,5} = E(X) = 2$]

4.33 Čas nepriameho bunkového delenia (nazývaný mitóza) má normálne rozdelenie so strednou hodnotou jedna hodina a smerodajnou odchýlkou 5 minút.

- a) Aká je pravdepodobnosť, že delenie nejakej bunky sa uskutoční skôr ako za 45 minút?
- b) Aká je pravdepodobnosť, že delenie nejakej bunky bude trvať viac ako 65 minút?
- c) Aký čas je potrebný na dokončenie mitózy približne 99 % zo všetkých buniek?

[a)0,00135; b)0,15866; c)71,6 min]

4.34 Reakčná doba šoféra na zrkový podnet má normálne rozdelenie so strednou hodnotou 0,4 sekundy a smerodajnou odchýlkou 0,05 sekundy.

- a) Aká je pravdepodobnosť, že reakcia nejakého šoféra vyžaduje 0,5 sekundy?
- b) Aká je pravdepodobnosť, že reakcia nejakého šoféra je medzi 0,4 až 0,5 sekúnd?
- c) Aká je reakčná doba, ktorú prekročí 90 % šoférov?

[0,02275; 0,47785; 0,336]

4.35 Rýchlosť prenosu súboru zo serveru na osobný počítač v určitej sieti má normálne rozdelenie so strednou hodnotou 60 Mb/s a smerodajnou odchýlkou 4 Mb/s.

- a) Aká je pravdepodobnosť, že súbor bude prenesený väčšou rýchlosťou ako 70 Mb/s?
- b) Aká je pravdepodobnosť, že súbor bude prenesený rýchlosťou menšou ako 58 Mb/s?

[a) = $1 - \text{NORM.DIST}(70; 60; 4; 1) = 0,00621$; b) $\text{NORM.DIST}(58; 60; 4; 1) = 0,308538$]

4.36 Práceschopnosť zamestnancov vo firme za mesiac má normálne rozdelenie so strednou hodnotou 100 hodín a smerodajnou odchýlkou 18 hodín. Aká je pravdepodobnosť, že práceschopnosť na budúci mesiac bude 80 až 90 hodín? Aký veľký čas by mal byť zahrnutý do rozpočtu na práceschopnosť, ak by plánované absencie mali byť prekročené s pravdepodobnosťou iba 10 %?

[= $\text{NORM.DIST}(90; 100; 18; 1) - \text{NORM.DIST}(80; 100; 18; 1) = 0,155997$;
= $\text{NORM.INV}(0,9; 100; 18) = 123,0679$ hodín]

4.37 Hmotnosť výrobku je vyhovujúca, ak je v medziach 59-61 gramov. Za štandardných podmienok má hmotnosť výrobkov približne normálne rozdelenie so strednou hodnotou $\mu = 60,05$ gramov a smerodajnou odchýlkou $\rho = 0,7$. Aká je pravdepodobnosť, že hmotnosť výrobku bude vyhovujúca?

[= $\text{NORM.DIST}(61; 60,05; 0,7; 1) - \text{NORM.DIST}(59; 60,05; 0,7; 1) = 0,845825$]

4.38 Pevnosť vzoriek kontajnerov cementu v tlaku sa modeluje pomocou normálneho rozdelenia so strednou hodnotou 6000 kg/cm^2 a smerodajnou odchýlkou 100 kg/cm^2 .

- a) Aká je pravdepodobnosť, že pevnosť nejakej vzorky je menšia ako 6250 kg/cm^2 ?
- b) Aká je pravdepodobnosť, že pevnosť nejakej vzorky je medzi 5800 a 5900 kg/cm^2 ?
- c) Akú pevnosť prekročí 95 % vzoriek? [a)0,99379; b)0,13591; c)5835 kg/cm^2]

4.39 Na obrábacom stroji sa sleduje dĺžka konkrétnej pracovnej operácie, ktorá má normálne rozdelenie so strednou hodnotou 21 minút a smerodajnou odchýlkou 1,3 minúty. Aká je pravdepodobnosť, že náhodne vybraná pracovná operácia neprekročí 20 minút? [0,221]

4.40 Doba ošetrovania pacienta u lekára je náhodná veličina, u ktorej je možné predpokladať normálne rozdelenie so strednou hodnotou 14 minút a smerodajnou odchýlkou 5 minút. Aká je pravdepodobnosť, že:

- a) náhodne vybraný pacient strávi na vyšetrení viac ako 20 minút?
- b) celková doba vyšetrenia u 24 objednaných pacientov prekročí 6 hodín?

[a) 0,115; b) 0,164]

4.41 Meranie výrobku je sprevádzané náhodnými chybami, ktoré majú smerodajnú odchýlku 4 mm. Určte pravdepodobnosť, že chyba merania neprekročí v absolútnej hodnote 2 mm.
[0,383]

4.42 Predpokladajme, že telesná hmotnosť v danej populácii sa riadi normálnym rozdelením, priemerná hmotnosť je 80 kg a smerodajná odchýlka je 10 kg.

- a) Určte, koľko percent jedincov má hmotnosť najviac 105 kg.
- b) Určte, s akou pravdepodobnosťou bude hmotnosť náhodne vybraného jedinca z danej populácie najviac 75 kg.
- c) Určte, s akou pravdepodobnosťou bude hmotnosť náhodne vybraného jedinca z danej populácie aspoň 75 kg.
- d) Určte, s akou pravdepodobnosťou bude hmotnosť náhodne vybraného jedinca z danej populácie v rozmedzí 75 kg až 105 kg.
- e) Stanovte hmotnostnú hranicu, ktorú prekročí iba 5% jedincov.

[a)0,994 ; b)0,3085; c)0,6915; d)0,6853;e) 96,45 kg]

4.43 Priemer hriadeľa mechaniky optickej pamäte má normálne rozdelenie so strednou hodnotou 0,637 cm a smerodajnou odchýlkou 0,0013 cm. Špecifikácie na hriadeľ sú $0,6350 \pm 0,0039$ cm. Aký podiel hriadeľov vyhovuje špecifikáciám?

[=NORM.DIST(0,635+0,0039;0,637;0,0013;1)-NORM.DIST(0,635-0,0039;0,637;0,0013;1) = 0,928063]

4.44 Meracím prístrojom, ktorý má systematickú chybu +0,1; sa merala určitá fyzikálna veličina. Náhodné chyby meraní majú rozdelenie $N(0; 0,05)$.

Aká je pravdepodobnosť, že chyba meraní v absolútnej hodnote neprekročí 0,5?

[=NORM.DIST(0,5;0,1;SQRT(0,05);1)-NORM.DIST(-0,5;0,1;SQRT(0,05);1) = 0,959536]

4.45 Výrobok je v norme, ak je jeho dĺžka z intervalu 68 až 69 mm. Z dlhodobých pozorovaní je známe, že dĺžka daných výrobkov má normálne rozdelenie pravdepodobnosti so strednou hodnotou 68,2 a disperziou 0,04. Aká je pravdepodobnosť, že vybraný výrobok bude vyhovujúci?

[=NORM.DIST(69;68,2;0,2;1)-NORM.DIST(68;68,2;0,2;1) = 0,841313]

4.46 Hmotnosť vyrábaného závažia má normálne rozdelenie pravdepodobnosti so strednou hodnotou 20g, pričom výrobca uvádza jej štandardnú (smerodajnú) odchýlku 0,02g. Určte pravdepodobnosť toho, že náhodne kúpené závažie bude mať skutočnú hmotnosť:

- a) aspoň 20,03g;
- b) menšiu ako 19,99g;

c) aspoň 20g a nie viac ako 20,5g.

[a) = 1 – NORM. DIST(20,03; 20; 0,02; 1) = 0,066807; b) = NORM. DIST(19,99; 20; 0,02; 1) = 0,308538 ; c) = NORM. DIST(20,05; 20; 0,02; 1) – NORM. DIST(20; 20; 0,02; 1) = 0,49379]

4.47 Dĺžka distribuovaných hriadeľov má normálne rozdelenie so strednou hodnotou 45 cm a rozptylom 0,25 cm. Aká je pravdepodobnosť, že náhodne vybraný hriadeľ bude:

- a) kratší ako 45,25 cm;
- b) dlhší ako 45,25 cm;
- c) mať dĺžku od 45 cm do 45,25 cm.

[a) = NORM. DIST(45,25; 45; 0,5; 1) = 0,691462; b) = 1 – NORM. DIST(45,25; 45; 0,5; 1) = 0,308538; c) = 0,691462 – 0,5 = 0,191462]

4.48 Hmotnosť vyrábaných súčiastok je normálne rozdelená veličina so strednou hodnotou 110 gramov a rozptylom 100. S akou pravdepodobnosťou bude hmotnosť súčiastky menšia ako 115 gramov?

[= NORM. DIST(115; 110; 10; 1) = 0,691462]

4.2.3 Exponenciálne rozdelenie (exponential distribution)

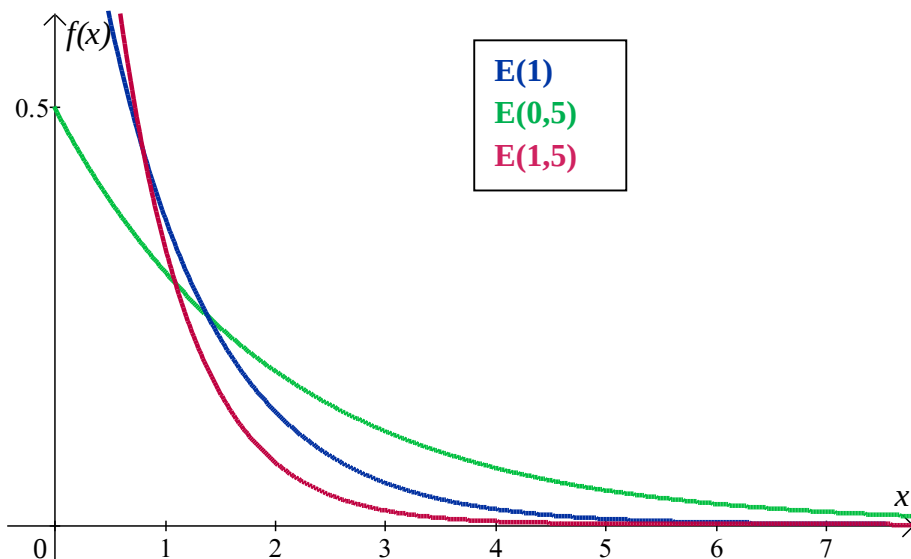
Exponenciálne rozdelenie pravdepodobnosti spojitej náhodnej premennej je obdobou Poissonovho rozdelenia pre diskretnú náhodnú premennú. Máme tok udalostí s intenzitou λ udalostí za jednotku času. Spojitá náhodná premenná X predstavuje časový interval medzi dvoma po sebe idúcimi javmi. Je zrejmé, že X nadobúda len nezáporné hodnoty (čas $t \geq 0$). Funkcia hustoty pravdepodobnosti je funkciou dĺžky časového intervalu t medzi dvoma po sebe nasledujúcimi výskytmi náhodných javov.

Hustota a distribučná funkcia:

$X \sim E(\lambda)$		
$f(x)$	$= \begin{cases} \lambda e^{-\lambda x} & \text{pre } x \geq 0 \\ 0 & \text{pre } x < 0 \end{cases}$	$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{pre } x \geq 0 \\ 0 & \text{pre } x < 0 \end{cases}$
Excel:	=EXPON.DIST(x;lambda;kumulatívne) x je hodnota náhodnej premennej X ; lambda je strednou hodnotou rozdelenia; kumulatívne pre 0 vypočíta hodnotu $f(x)$, pre 1 hodnotu $F(x)$.	

Číselné charakteristiky:

$E(X)$	$D(X)$	$\alpha(X)$	$\beta(X)$
$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2	6



Obr. 4.23 Grafy funkcie hustoty pravdepodobnosti pre vybrané hodnoty parametra λ

Toto rozdelenie dobre opisuje rozdelenie pravdepodobnosti doby životnosti zariadenia, u ktorého prichádza k poruche z náhodných príčin a nie v dôsledku mechanického opotrebovania alebo únavy materiálu (životnosť je uvažovaná ako doba „čakania na poruchu“). Používa sa v biológii v analýze doby prežitia, vo fyzike pri modelovaní rýchlosti rozpadu izotopov. Tiež sa často používa v teórii spoľahlivosti, v teórii hromadnej obsluhy, v teórii obnovy.

Toto rozdelenie má náhodná premenná popisujúca:

- čas medzi dvoma prístupmi na server;
- čas do zlyhania (životnosť zariadení, určovanie záručnej lehoty na výrobok);
- čas potrebný na opravu pokazenej súčiastky.

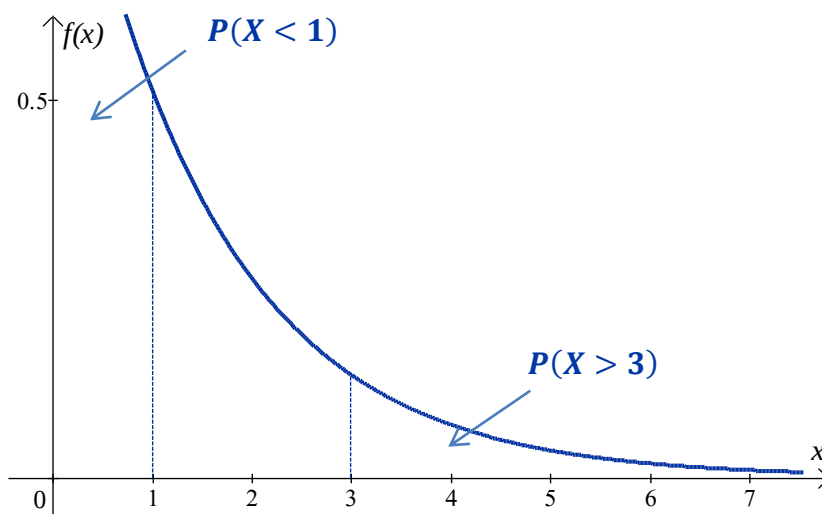
Poznámka 4.5 Exponenciálne a Poissonovo rozdelenie vyjadrujú svojím spôsobom tú istú podstatu z rôznych pohľadov. Kým pri Poissonovom rozdelení ide o pravdepodobnosť výskytu určitého počtu javov za jednotku času, pri exponenciálnom rozdelení nám ide o pravdepodobnosť dĺžky intervalu medzi dvoma výskytmi javu. Čím je časový interval medzi výskytmi dvoch javov dlhší, tým je pravdepodobnosť výskytu tohto intervalu nižšia.

Exponenciálne rozdelenie je špeciálnym prípadom gama rozdelenia.

Príklad 4.12 Stredná doba čakania zákazníka na obsluhu v predajni je 1,5 minúty. Predpokladajme, že náhodná premenná X , vyjadrujúca dobu čakania na obsluhu, má exponenciálne rozdelenie. Aká je pravdepodobnosť, že náhodný zákazník bude obslužený v dobe kratšej ako 1 minúta? Aký je podiel zákazníkov, ktorí budú čakať dlhšie ako 3 minúty?

Riešenie. Predpokladáme, že $X \sim E(\lambda)$, kde X je doba čakania na obsluhu.

Stredná hodnota tohto rozdelenia $E(X) = \frac{1}{\lambda} = 1,5 \Rightarrow \lambda = 0, \bar{6}$.



Obr. 4.24 Grafy funkcie hustoty pravdepodobnosti $X \sim E(0, \bar{6})$

$$P(X < 1) = F(1) = 1 - e^{-0,6 \cdot 1} = 0,486582881 ;$$

Hodnotu distribučnej funkcie v Exceli vypočítame pomocou =EXPON.DIST(1;1/1,5;1).

Zákazník bude obslužený v dobe kratšej ako 1 minúta s pravdepodobnosťou 48,66%.

$$P(X > 3) = 1 - F(3) = e^{-0,6 \cdot 3} = 0,135335283 ;$$

Výpočet v Exceli: =1-EXPON.DIST(3;1/1,5;1)

13,53% zákazníkov bude čakať na obsluhu dlhšie ako 3 minúty.

Úlohy

4.49 Priemerná životnosťou LED diód je 40 000 hodín. Za predpokladu, že pre modelovanie životnosti LED diódy (doby do poruchy) použijete exponenciálne rozdelenie, určte pravdepodobnosť toho, že LED dióda vydrží v prevádzke:

- aspoň 30 000 hodín,
- najviac 35 000 hodín,
- 30 000 – 50 000 hodín.

$$\left[X \sim E\left(\frac{1}{40\,000}\right); \text{ a) } P(X \geq 30000) = 1 - \text{EXPON.DIST}\left(30000; \frac{1}{40000}; 1\right) \rightarrow 0,472367; \text{ b) } P(X \leq 35000) = \text{EXPON.DIST}(35000; 1/40000; 1) \rightarrow 0,583138; \text{ c) } = \text{EXPON.DIST}(50000; 1/40000; 1) - \text{EXPON.DIST}(30000; 1/40000; 1) \rightarrow 0,185862 \right]$$

4.50 Doba životnosti výrobku má exponenciálne rozdelenie pravdepodobnosti so strednou hodnotou 60 000 hodín. Určte pravdepodobnosť toho, že výrobok nebude funkčný dlhšie, ako je jeho priemerná doba životnosti. Akú maximálnu dobu na výmenu nefunkčného výrobku za nový môže výrobca poskytnúť, ak pripúšťa 15% reklamačných výrobkov?

$$\left[X \sim E\left(\frac{1}{60\,000}\right); P(X \leq 60\,000) = \text{EXPON.DIST}\left(60\,000; \frac{1}{60\,000}; 1\right) \rightarrow 0,632121; = \right. \\ \left. \text{EXPON.DIST}\left(50\,000; \frac{1}{40\,000}; 1\right) - \text{EXPON.DIST}\left(30\,000; \frac{1}{40\,000}; 1\right) \rightarrow 0,185862; P(X \leq x) = \right. \\ \left. F(x) = 1 - e^{-\frac{x}{60\,000}} = 0,15 \rightarrow x = 9\,751,14 \right]$$

4.51 Určitá súčiastka sa pokazí v záručnej dobe 1095 dní s pravdepodobnosťou 0,1. Je známe, že doba životnosti takejto súčiastky sa riadi exponenciálnym rozdelením pravdepodobnosti. Určte :

- strednú dobu životnosti tejto súčiastky;
- pravdepodobnosť toho, že súčiastka sa nepokazí počas jedného roka (365 dní);
- pravdepodobnosť toho, že súčiastka sa nepokazí počas prvých 200 dní, ale pokazí sa v priebehu ďalších 500 dní.

4.2.4 Weibullovo rozdelenie (Weibull distribution)

Toto rozdelenie vyjadruje životnosť, bezporuchovosť technických zariadení (tzv. dobu do poruchy), ktorým nevyhovuje exponenciálne rozdelenie, t.j. kde sa prejavuje mechanické opotrebovanie a únava materiálu. Všeobecne môžeme týmto rozdelením modelovať dátové súbory, ktorých hodnoty sú väčšie ako nula (pri analýzach životnosti a bezporuchovosti priemyselných systémov a výrobkov v elektronickom, elektrotechnickom, strojárskom priemysle).

Funkcia hustoty a distribučná funkcia:

$X \sim W(\lambda, k)$			
$f(x)$	$= \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} & \text{pre } x \geq 0 \\ 0 & \text{pre } x < 0 \end{cases}$	$F(x)$	$= \begin{cases} 1 - e^{-\left(\frac{x}{\lambda}\right)^k} & \text{pre } x \geq 0 \\ 0 & \text{pre } x < 0 \end{cases}$

Číselné charakteristiky:

$E(X)$	$D(X)$	$\alpha(X)$	$\beta(X)$
$\lambda \Gamma\left(1 + \frac{1}{k}\right)$	$\lambda^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2 \right]$		

Príklady Weibullovo rozdelenia:

Toto rozdelenie sa používa:

- v prípadoch, kedy bezporuchovosť závisí na veku, počte odpracovaných hodín, počte vykonaných cyklov, opotrebení (popisuje čas do poruchy systému, kedy porucha nastáva pri poruche najslabšieho komponentu systému);
- v oblasti spoľahlivosti pri určovaní ukazovateľov bezporuchovosti, ktoré predstavujú dôležitú informáciu pri predpovedi, hodnotení či porovnaní životnosti výrobkov, vyhodnotení konštrukčných a technologických zmien, porovnaní alternatívnych

konštrukcií a technológií, porovnaní životnosti výrobkov rôznych technológií alebo rôznych výrobcov, vytváraní záručnej politiky, pri proaktívnom prístupe k riadeniu zásob náhradných dielov alebo pri plánovaní opráv;

- pri modelovaní rôznych javov: napr. predpoveď počasia, úmrtnosť na AIDS, hodnotenie pravdepodobnosti vzniku zemetrasenia,...

4.3 ROZDELENIA FUNKCIÍ NÁHODNÝCH PREMENNÝCH (ODVODENÉ ROZDELENIA, VÝBEROVÉ ROZDELENIA)

Majú mimoriadny význam v praxi pri analýze štatistických údajov získaných náhodným výberom, t.j. používajú sa pri výberovom skúmaní (χ^2 rozdelenie pri popise výberového rozptylu, Studentovo t -rozdelenie najčastejšie k porovnaniu priemerov, Fisherovo F -rozdelenie pri porovnaní rozptylov dvoch súborov alebo pri porovnaní závislých veličín v lineárnej regresii). Sú úzko spojené s normálnym rozdelením.

4.3.1 χ^2 rozdelenie (chí-kvadrát, Helmertovo-Pearsonovo rozdelenie) (Chi-squared distribution)

Nech $X_1, X_2, X_3, \dots, X_n$ je n nezávislých náhodných premenných, z ktorých každá má normované normálne rozdelenie $N(0,1)$. Ich súčet štvorcov je náhodná premenná, ktorá má rozdelenie, ktoré nazývame chí-kvadrát rozdelenie s n stupňami voľnosti (degree of freedom), zapisujeme $\chi^2 \sim \chi^2(n)$. Označujeme

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2 = \sum_{i=1}^n X_i^2$$

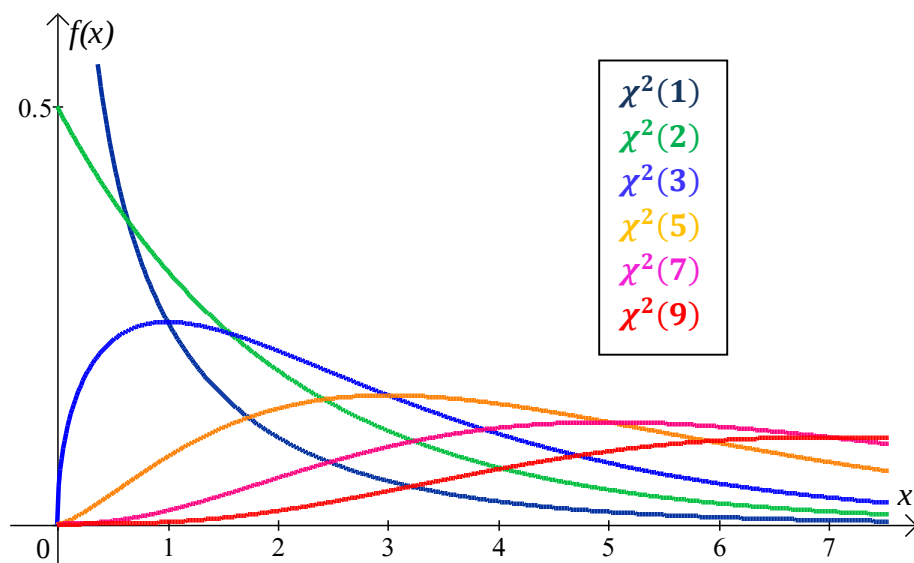
Funkcia hustoty:

$\chi^2 \sim \chi^2(n)$	
$f(x)$	$= \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}, x \in (0, \infty)$
Excel:	=CHISQ.DIST(x;stupeň voľnosti;kumulatívne) x je hodnota náhodnej premennej χ^2 ; stupeň voľnosti je parameter určujúci počet stupňov voľnosti; kumulatívne je logická premenná určujúca, či bude vypočítaná funkcia hustoty pravdepodobnosti (parameter 0) alebo distribučnú funkcia (parameter 1).

Číselné charakteristiky:

$E(\chi^2)$	$D(\chi^2)$	$\alpha(\chi^2)$	$\beta(\chi^2)$
n	$2n$	$\sqrt{\frac{8}{n}}$	$3 + \frac{12}{n}$

Grafy funkcie hustoty pravdepodobnosti chí-kvadrát rozdelenia pre vybrané hodnoty parametra n sú na obr. 4.25.



Obr. 4.25 Grafy funkcie hustoty pravdepodobnosti chí-kvadrát rozdelenia pre vybrané hodnoty parametra n

Kvantily chí-kvadrát rozdelenia

Kvantily chí-kvadrát rozdelenia budeme označovať $\chi_n^{2,kv}(\alpha)$, kde $P(\chi^2 \leq \chi_n^{2,kv}(\alpha)) = \alpha$. Na výpočet kvantilov chí-kvadrát rozdelenia sa používa kvantilová funkcia `=CHISQ.INV(pravdepodobnosť; stupeň voľnosti)`, kde pravdepodobnosť je pravdepodobnosť α , pre ktorú hľadáme kvantil $\chi_n^{2,kv}(\alpha)$; stupeň voľnosti je parameter, určujúci počet stupňov voľnosti.

Vlastnosti: Počet stupňov voľnosti n je daný počtom nezávislých sčítancov a je jediným parametrom rozdelenia. S rastúcim počtom stupňov voľnosti sa chí-kvadrát rozdelenie „blíži“ k normálnemu rozdeleniu, t.j. pre $n \geq 30$ ho možno aproximovať (nahradit) normálnym rozdelením $N(n; 2n)$.

$$n \geq 30 \Rightarrow \chi^2(n) \sim N(n; 2n).$$

4.3.2 Studentovo t -rozdelenie (Student's t -distribution)

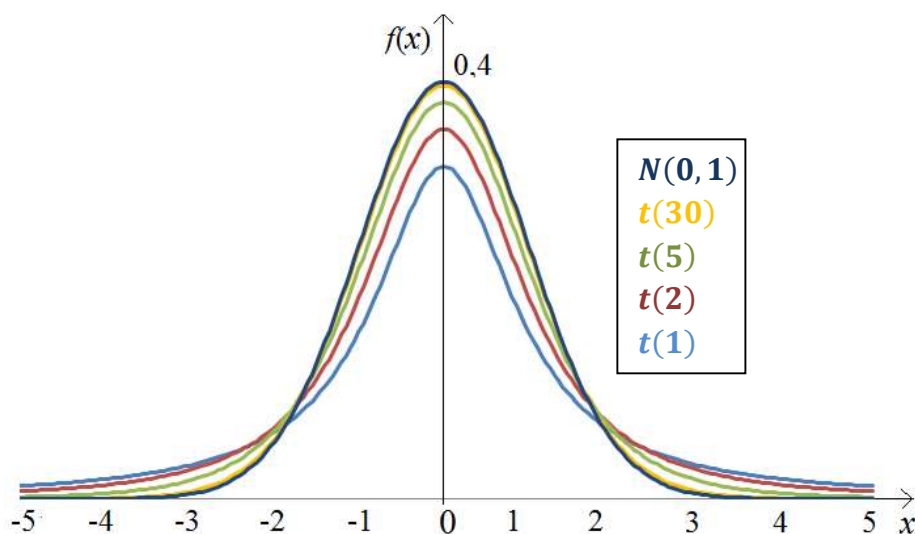
Nech $X, X_1, X_2, X_3, \dots, X_n$ je $n + 1$ nezávislých náhodných premenných, z ktorých každá má normované normálne rozdelenie $N(0,1)$. Označme:

$$t = \frac{X}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}} = \frac{X}{\sqrt{\frac{1}{n} \chi^2}}$$

Premenná t je náhodná premenná. Jej rozdelenie sa nazýva Studentovo t -rozdelenie s n stupňami voľnosti $t \sim t(n)$.

Vlastnosti: Pre každý výber s rozsahom $n + 1$ (t.j. počet stupňov voľnosti $df = n$) existuje odlišné t -rozdelenie, preto hovoríme, že počet stupňov voľnosti je jediný parameter rozdelenia. Funkcia hustoty rozdelenia pravdepodobnosti je párna – jej krivka je súmerná podľa y -ovej osi, podobná na krivku normálneho rozdelenia, len je plochejšia (má viac plochy na okrajoch a menej v strede). V porovnaní s normálnym rozdelením má vyššiu variabilitu, v dôsledku toho budú intervaly spoľahlivosti širšie ako u normálneho rozdelenia. S rastúcim počtom stupňov voľnosti sa t -rozdelenie „blíži“ k normovanému normálnemu rozdeleniu, pre $n \geq 30$ ho možno aproximovať (nahradit) normovaným normálnym rozdelením $N(0; 1)$, obr. 4.26.

$$n \geq 30 \Rightarrow t(n) \sim N(0; 1)$$



Obr. 4.26 Porovnanie grafu funkcie hustoty Studentovho t -rozdelenia pre rôzne stupne voľnosti s grafom Gaussovej krivky rozdelenia $N(0,1)$

Funkcia hustoty:

$t \sim t(n)$	
$f(x)$	$= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2}) \cdot \sqrt{\pi n}} \cdot \left(1 + \frac{1}{x^2}\right)^{-\frac{n+1}{2}}, x \in (-\infty, \infty)$
Excel:	Na výpočet funkcie hustoty a distribučnej funkcie sa používa $=T.DIST(x; \text{stupeň voľnosti}; \text{kumulatívne})$ x je hodnota náhodnej premennej; stupeň voľnosti je parameter určujúci počet stupňov voľnosti; kumulatívne je logická premenná určujúca, či bude vypočítaná funkcia hustoty pravdepodobnosti (parameter 0) alebo distribučnú funkcia (parameter 1)

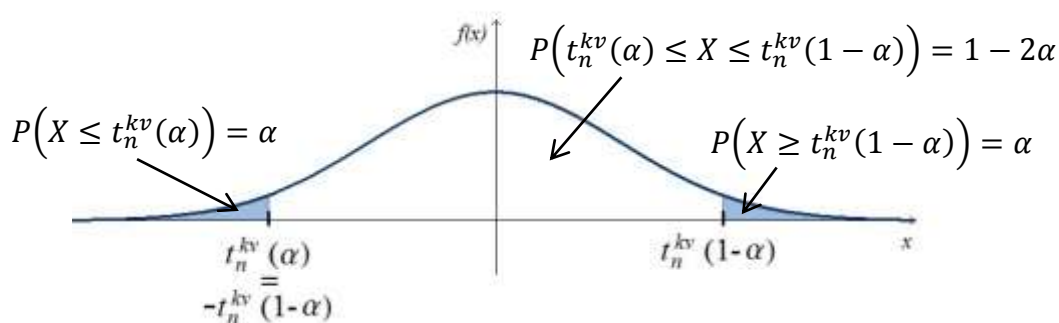
Číselné charakteristiky:

$E(t)$	$D(t)$	$\alpha(t)$	$\beta(t)$
0 pre $n > 1$	$\frac{n}{n-2}$ pre $n > 2$	0	$\frac{6}{n-4}$ pre $n > 4$

S týmto rozdelením sa stretávame v teórii odhadu, pri testoch štatistických hypotéz a pod.

Kvantily t -rozdelenia

Kvantily t -rozdelenia budeme označovať $t_n^{kv}(\alpha)$, kde $P(t \leq t_n^{kv}(\alpha)) = \alpha$, kde α je pravdepodobnosť „naľavo“ od kvantilu (veľkosť „zľava ohraničenej“ plochy pod hustotou rozdelenia). Na výpočet kvantilov rozdelenia v Exceli sa používa kvantilová funkcia $=T.INV(\text{pravdepodobnosť}; \text{stupeň voľnosti})$, pravdepodobnosť je pravdepodobnosť α , pre ktorú hľadáme kvantil $t_n^{kv}(\alpha)$; stupeň voľnosti je parameter, určujúci počet stupňov voľnosti. Zo symetrie Studentovho t -rozdelenia vyplýva, že $t_n^{kv}(\alpha) = -t_n^{kv}(1 - \alpha)$, obr. 4.27.

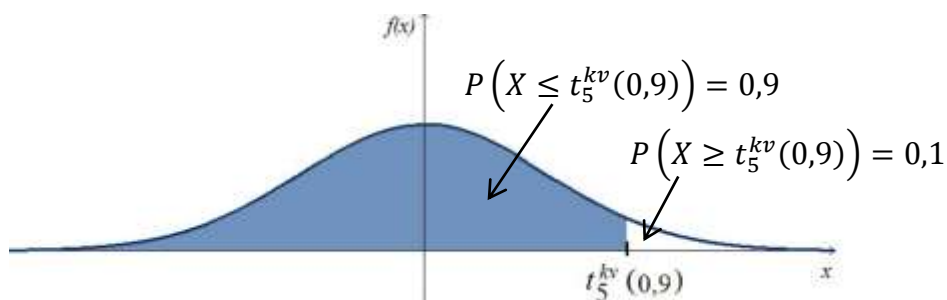


Obr. 4.27 Z tvaru Studentovho t -rozdelenia vyplýva, že $t_n^{kv}(\alpha) = -t_n^{kv}(1 - \alpha)$

Príklad 13. Vypočítajte 90% kvantil Studentovho t -rozdelenia s 5 stupňami voľnosti. Aký je podiel hodnôt náhodnej premennej s daným rozdelením, ktoré presahujú číselnú hodnotu vypočítaného kvantilu?

Riešenie. $t_5^{kv}(0,9) = T.INV(0,9; 5) = 1,475884$

10% hodnôt náhodnej premennej so Studentovým t -rozdelením s 5 stupňami voľnosti presahuje číselnú hodnotu 1,475884, obr. 4.28.



Obr. 4.28 Zobrazenie kvantilu $t_5^{kv}(0,9)$ – jeho hodnota rozdeľuje veľkosť plochy pod krivkou na plochu s obsahom 0,9 a plochu s obsahom 0,1.

Úloha

4.52 Určte 95. a 99. percentil Studentovho t -rozdelenia so 4 stupňami voľnosti.

[$= T.INV(0,95; 4) = 2,131847$; $= T.INV(0,99; 4) = 3,746947$]

4.3.3 *F*-rozdelenie alebo Fisherovo – Snedecorovo rozdelenie (Snedecor's *F*-distribution or Fisher-Snedecor distribution)

Nech $X_1, X_2, X_3, \dots, X_{n_1}$ a $Y_1, Y_2, Y_3, \dots, Y_{n_2}$ sú nezávislé náhodné premenné, z ktorých každá má normované normálne rozdelenie $N(0,1)$. Potom rozdelenie náhodnej premennej

$$F = \frac{\frac{1}{n_1} \sum_{i=1}^{n_1} X_i^2}{\frac{1}{n_2} \sum_{i=1}^{n_2} Y_i^2} = \frac{\frac{1}{n_1} \chi_1^2}{\frac{1}{n_2} \chi_2^2}$$

sa nazýva Fisherovo – Snedecorovo rozdelenie s n_1 a n_2 stupňami voľnosti, označuje sa $F \sim F(n_1, n_2)$.

Vlastnosti: Fisherovo rozdelenie má 2 parametre, stupne voľnosti n_1 a n_2 , je asymetrické. Pre strednú hodnotu náhodnej premennej F platí $E(F) = \frac{n_2}{n_2 - 2}$ pre $n_2 > 2$. Na obr. 4.29 sú zobrazené funkcie hustoty pravdepodobnosti pre rôzne stupne voľnosti n_1, n_2 .

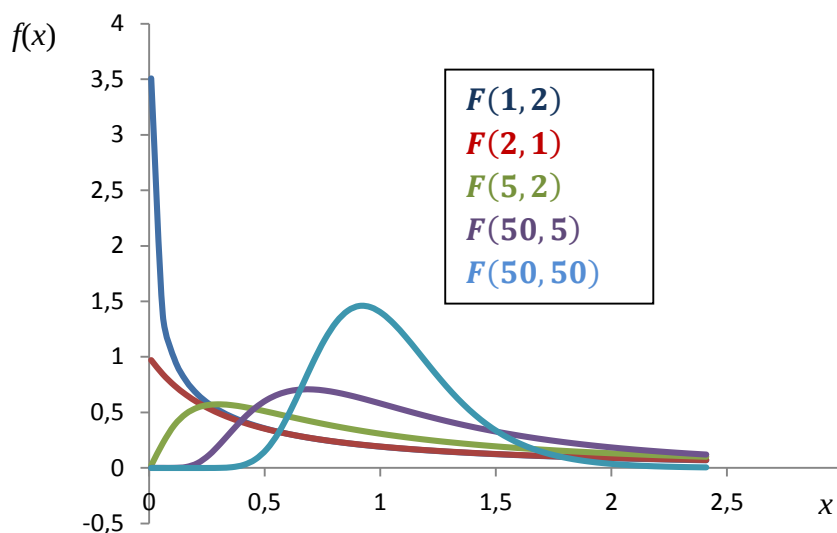
Distribučná funkcia a hustota rozdelenia v Exceli:

$F \sim F(n_1, n_2)$	
Excel:	Na výpočet funkcie hustoty a distribučnej funkcie sa používa =F.DIST(x; stupeň voľnosti_1; stupeň voľnosti_2;kumulatívne) x je hodnota náhodnej premennej; stupeň voľnosti_1, stupeň voľnosti_2 sú parametre určujúce počet stupňov voľnosti n_1, n_2 ; kumulatívne je logická premenná určujúca, či bude vypočítaná funkcia hustoty pravdepodobnosti (parameter 0) alebo distribučná funkcia (parameter 1).

Číselné charakteristiky:

$E(F)$	$D(F)$	$\alpha(F)$
$\frac{n_2}{n_2 - 2}$ pre $n_2 > 2$	$\frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$ pre $n_2 > 4$	$\frac{(2n_1 + n_2 - 2)\sqrt{8(n_2 - 4)}}{(n_2 - 6)\sqrt{n_1(n_1 + n_2 - 2)}}$ pre $n_2 > 4$

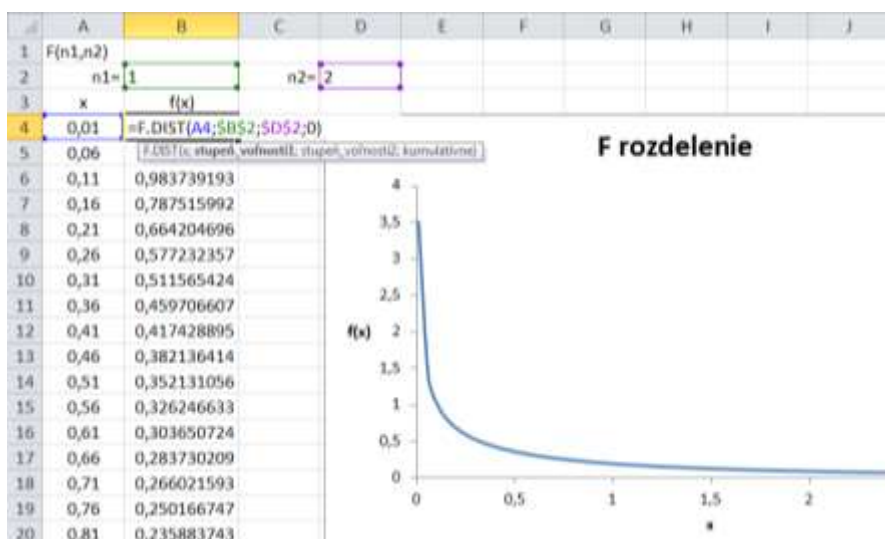
Porovnanie grafov funkcie hustoty F -rozdelenia pre vybrané stupne voľnosti n_1, n_2 môžeme vidieť na nasledujúcom obrázku.



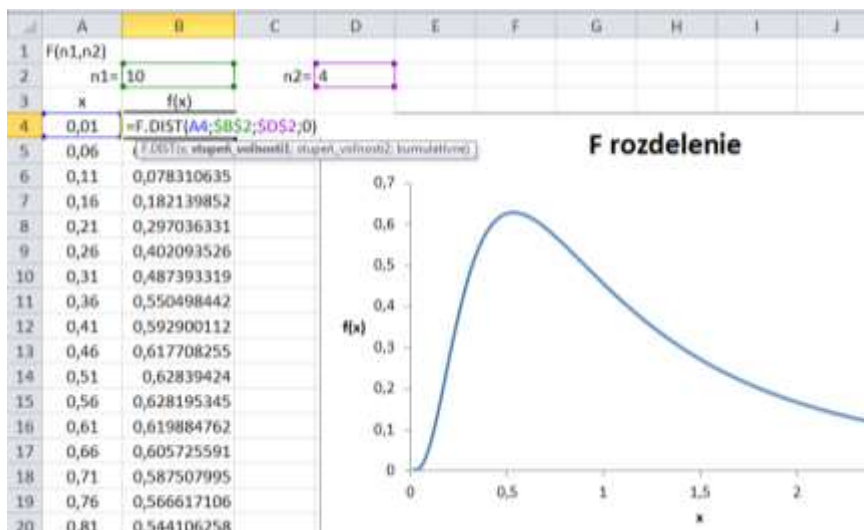
Obr. 4.29 Porovnanie grafov funkcie hustoty F -rozdelenia pre vybrané stupne voľnosti n_1, n_2

Príklad 4.14 Zobrazte pomocou Excelu graf funkcie hustoty F -rozdelenia a pozorujte, ako sa mení tvar krivky v závislosti od stupňov voľnosti n_1, n_2 .

Riešenie. Na zobrazenie funkcie hustoty náhodnej premennej, ktorá má F -rozdelenie, použijeme v Exceli preddefinovanú funkciu F.DIST. Jej použitie je naznačené na nasledujúcich obr. 4.30 a 4.31.



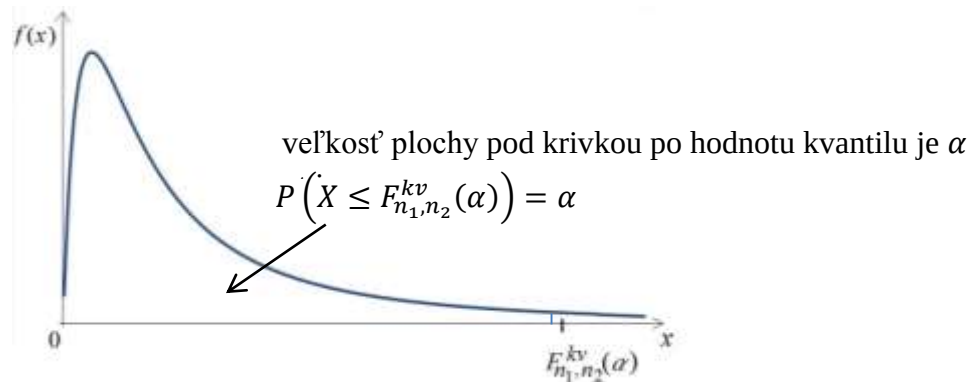
Obr. 4.30 Zobrazenie grafu funkcie hustoty pravdepodobnosti náhodnej premennej $F \sim F(1, 2)$ v Exceli



Obr. 4.31 Zobrazenie grafu funkcie hustoty pravdepodobnosti náhodnej premennej $F \sim F(10, 4)$ v Exceli .

Kvantily F -rozdelenia

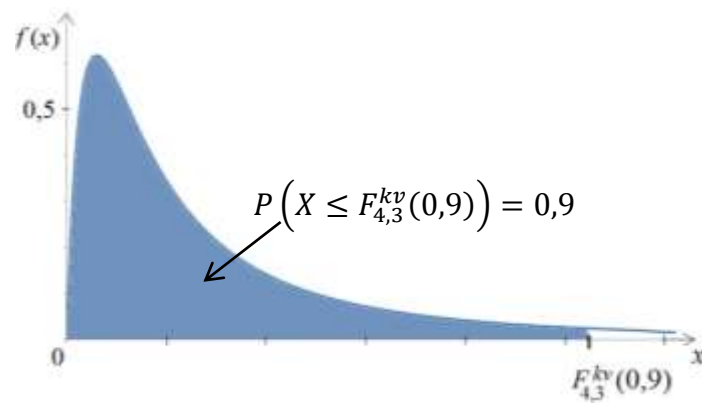
Kvantily F -rozdelenia budeme označovať $F_{n_1, n_2}^{kv}(\alpha)$, kde $P(F \leq F_{n_1, n_2}^{kv}(\alpha)) = \alpha$, kde α je pravdepodobnosť „naľavo“ od kvantilu (veľkosť „zľava ohraničenej“ plochy pod hustotou rozdelenia), obr. 4.32 . Na výpočet kvantilov rozdelenia v Exceli sa používa kvantilová funkcia $=F.INV(\text{pravdepodobnosť}; \text{stupeň_voľnosti 1}; \text{stupeň_voľnosti 2})$.



Obr. 4.32 Geometrická interpretácia kvantilu $F_{n_1, n_2}^{kv}(\alpha)$

Príklad 4.15 Nájdite 90. percentil náhodnej premennej s rozdelením $F \sim F(4, 3)$.

Riešenie. Na výpočet hodnoty kvantilu náhodnej premennej, ktorá má F -rozdelenie použijeme v Exceli funkciu F.INV, t.j. $F_{4,3}^{kv}(0,9) = F.INV(0,9; 4; 3) = 5,342644$. Toto číslo rozdeľuje plochu pod krivkou hustoty $f(x)$ tak, že 90% sa nachádza vľavo a 10% vpravo od tohto čísla (obr.4.33).



Obr. 4.33 Kvantil $F_{4,3}^{kv}(0,9)$

Úloha

4.53 Nech náhodná premenná X má F -rozdelenie $F(3, 5)$.

a) Nájdite 95. a 99. percentil tejto náhodnej premennej.

b) Nájdite číslo a tak, aby platilo $P(F \geq a) = 0,4$.

[a) $F_{3,5}^{kv}(0,95) = \text{F.INV}(0,95;3;5) = 5,409451$; $F_{3,5}^{kv}(0,99) = \text{F.INV}(0,99;3;5) = 12,05995$;

b) $a = F_{3,5}^{kv}(0,6) = \text{F.INV}(0,6;3;5) = 1,197805$]

5. VYUŽITIE ANALYTICKÝCH NÁSTROJOV A VYBRANÝCH ŠTATISTICKÝCH FUNKCIÍ V EXCELI

5.1 OPISNÁ ŠTATISTIKA

Definícia 5.1 Štatistická jednotka je základný prvok, na ktorom skúmame konkrétny prejav určitého hromadného javu.

Definícia 5.2 Štatistický znak je vonkajší, postihnuteľný, merateľný prejav vlastností štatistickej jednotky.

Definícia 5.3 Štatistický súbor je množina všetkých štatistických jednotiek.

Definícia 5.4 Základný štatistický súbor (populácia) je súbor všetkých štatistických jednotiek. Označujeme ZS.

Definícia 5.5 Výberový štatistický súbor (vzorka) je množina vybraných štatistických jednotiek. Používame označenie VS.

Druhy štatistického skúmania :

- **vyčerpávajúce** skúmanie - skúma sa celý základný súbor (každý štatistická jednotka),
- **výberové** skúmanie - skúma sa iba výberový súbor (vybrané štatistické jednotky).

Etapy štatistického skúmania: štatistické **zisťovanie** (meranie), štatistické **spracovanie** (usporiadanie, triedenie, výpočet číselných charakteristík) a **štatistická analýza**.

Parametre štatistického súboru:

základného súboru (označujú sa veľkými písmenami abecedy)

- M – stredná hodnota;
- σ^2 – smerodajná odchýlka;
- Q – ľubovoľný parameter;

výberového súboru (označujú sa malými písmenami abecedy):

- $x_1, x_2, \dots, x_n$ namerané hodnoty, n je rozsah výberového súboru;
- $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ – aritmetický priemer (stredná hodnota);
- $s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ – rozptyl výberového súboru.

Odhady parametrov ZS:

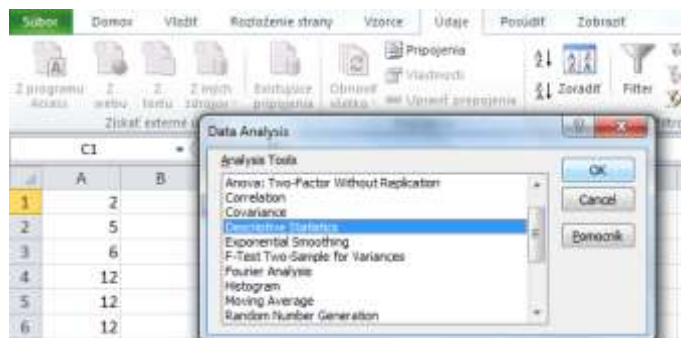
Bodový odhad (estimátor) parametra základného súboru – odhadneme jedným číslom (bodom) vypočítaným z VS (pomocou výberových charakteristík) – $estM = \bar{x}$, $est\sigma^2 = s_{n-1}^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$.

Intervalový odhad parametra Q základného súboru robíme s pravdepodobnosťou $1 - \alpha$. Zapisujeme $P(q_1 < Q < q_2) = 1 - \alpha$, kde $1 - \alpha$ sa nazýva spoľahlivosť intervalového

odhadu; α sa nazýva riziko intervalového odhadu. Interval (q_1, q_2) sa nazýva interval spoľahlivosti (konfidenčný interval).

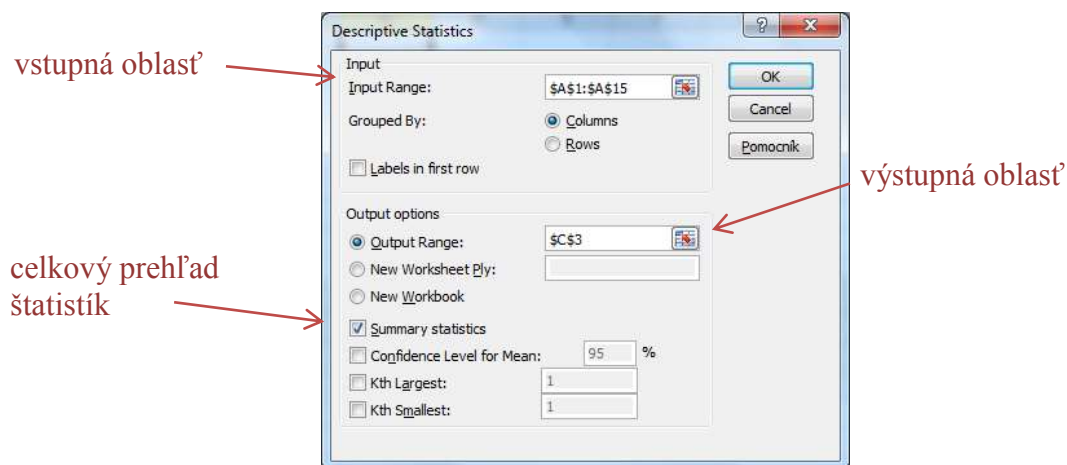
Príklad 5.1 Pomocou nástroja Opisná štatistika (Descriptive Statistics) urobte analýzu údajov výberového súboru VS: 2, 5, 6, 12, 12, 12, 20, 21, 22, 22, 23, 23, 25, 25, 26. Odhadnite parametre základného súboru ZS.

Riešenie. V rámci ponuky analytických nástrojov volíme nástroj Opisná štatistika (Descriptive Statistics) obr. 5.1, ktorý poskytuje súhrnný výpis vybraných charakteristík.



Obr. 5.1 Aktivácia jednotlivých analytických nástrojov sa realizuje postupnosťou príkazov Údaje → Analýza dát (Data → Data analysis)

Dialógové okno (obr. 5.2) požaduje vyplniť vstupnú oblasť (Input Range:), v ktorej špecifikujeme oblasť, v ktorej sa nachádzajú vstupné namerané údaje, v časti Grouped By určujeme informáciu, či sú vstupné údaje usporiadané v stĺpcoch (Columns) alebo riadkoch (Rows), ak sa v prvom riadku nachádza názov, popis, označíme voľbu Labels in first row, v časti výstupných podmienok (Output options) volíme špecifikáciu umiestnenia začiatku výstupu do bunky (Output Range:), do nového hárku (New Worksheet Ply:) alebo do nového súboru (New Workbook). V ďalšej časti volíme, ktoré štatistiky požadujeme vo výstupe: prehľad opisných štatistík (Summary statistics), prípustnú chybu odchýlky pre zadaný (obojsmerný) interval spoľahlivosti pre strednú hodnotu (Confidence Level for Mean), K-tu najväčšiu a K-tu najmenšiu hodnotu (Kth Largest, Kth Smallest).



Obr. 5.2 Dialógové okno analytického nástroja Popisná štatistika

Po výbere a potvrdení požadovaných štatistík je vytvorená tabuľka, ktorá obsahuje zvolené výstupy. Analýzou údajov zo zadania príkladu dostaneme nasledujúce výsledky, obr. 5.3.

	A	B	C	D
1	2		Príklad. Popisná štatistika	
2	5			
3	6		Column1	
4	12			
5	12		Mean	17,06667
6	12		Standard Error	2,094134
7	20		Median	21
8	21		Mode	12
9	22		Standard Deviation	8,110546
10	22		Sample Variance	65,78095
11	23		Kurtosis	-1,02712
12	23		Skewness	-0,69009
13	25		Range	24
14	25		Minimum	2
15	26		Maximum	26
16			Sum	256
17			Count	15

Obr. 5.3 Výstupná tabuľka analytického nástroja Popisná štatistika

Výstupná tabuľka však obsahuje niekoľko parametrov popisujúcich VS a niekoľko parametrov popisujúcich ZS. Všetky požadované hodnoty je možné dopočítať pomocou definičných vzťahov alebo pomocou preddefinovaných funkcií, napr. funkcia AVERAGE počíta aritmetický priemer, MODE modus, MEDIAN medián, STDEV.S smerodajnú odchýlku, VAR.S rozptyl, KURT koeficient špicatosti, SKEW koeficient šikmosti, ...

Charakteristiky (parametre) výberového súboru VS sú:

rozsah súboru $n = 15$ (Count), súčet všetkých hodnôt štatistického znaku vo VS = 256(Sum).

Charakteristiky strednej hodnoty VS: aritmetický priemer $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 17,06667$ (Mean), medián $Me = 21$ (Median), modus $Mo = 12$ (Mode).

Charakteristiky variability VS: maximálna hodnota $x_{max} = 26$ (Maximum), minimálna hodnota $x_{min} = 2$ (Minimum), rozpätie $R = x_{max} - x_{min} = 24$ (Range), rozptyl (tzv. vychýlený rozptyl) $s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 61,39556$ nie je súčasťou tabuľky, musíme ho dopočítať pomocou funkcie =VAR.P(číslo1;číslo2;...), smerodajnú odchýlku $s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = 7,835532$ vypočíta funkcia =STDEV.P(číslo1;číslo2;...).

Tvar rozdelenia popisujú (tzv. štandardizovaný) koeficient špicatosti $\beta = -1,02712$ (Kurtosis) a (tzv. štandardizovaný) koeficient šikmosti $\alpha = -0,69009$ (Skewness).

Poznámka 5.1 Súčasťou výstupu Popisnej štatistiky je aj chyba strednej hodnoty (Standard error), ktorá je vypočítaná ako podiel smerodajnej odchýlky a odmocniny rozsahu súboru.

Bodové odhady parametrov základného súboru ZS

Bodový odhad strednej hodnoty M ZS označujeme $est M = \bar{x} = 2,094134$, nazývame výberový priemer. Na jeho určenie môžeme použiť výsledné údaje z výstupu Opisnej štatistiky z Analýzy dát alebo použiť funkciu =AVERAGE(číslo1;číslo2;...).

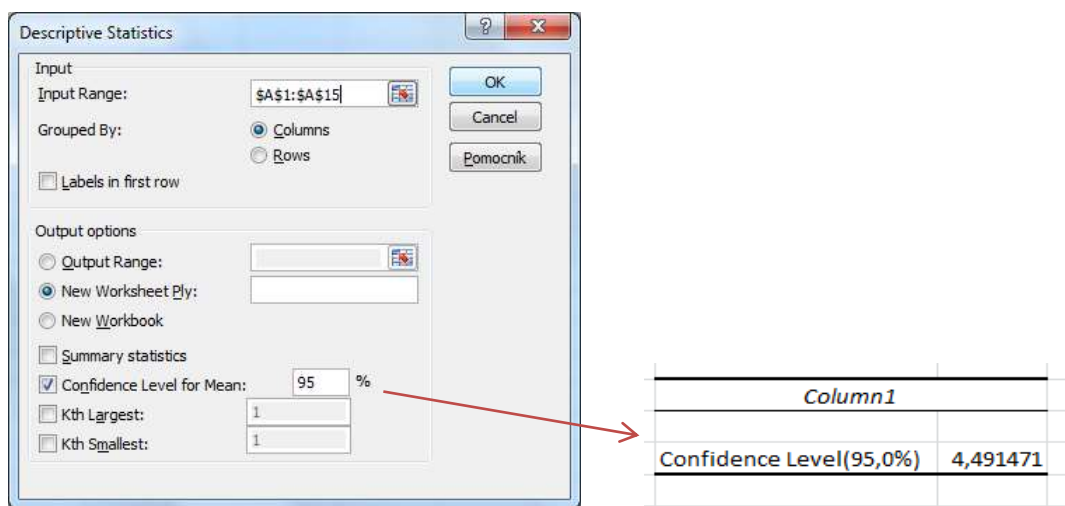
Bodový odhad rozptylu σ^2 ZS označujeme $est \sigma^2 = s_{n-1}^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 = 65,78095$, nazývame výberový rozptyl (nevychýlený rozptyl). Jeho číselná hodnota je výstupom Opisnej štatistiky z Analýzy dát, resp. použijeme funkciu =VAR.S(číslo1;číslo2;...). Medzi vychýleným a nevychýleným rozptylom platí vzťah $s_{n-1}^2 = \frac{n}{n-1} s_n^2$.

Bodový odhad smerodajnej odchýlky σ označujeme $est \sigma = s_{n-1} = \sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2} = 8,110546$ nazývame výberová smerodajná odchýlka. Jeho číselná hodnota je výstupom Opisnej štatistiky z Analýzy dát, resp. možno použiť funkciu =STDEV.S(číslo1;číslo2;...).

Intervalové odhady parametrov základného súboru ZS

Pre intervalový odhad strednej hodnoty M ZS platí $P(\bar{x} - \Delta < M < \bar{x} + \Delta) = 1 - \alpha$.

Excel umožňuje priamy výpočet prípustnej chyby odchýlky Δ pomocou funkcie =CONFIDENCE.T(alfa;smerodajná_odch;veľkosť) alebo pomocou analytického nástroja Popisnej štatistiky a voľbou konfidenčného intervalu so zadáním požadovanej spoľahlivosti $1 - \alpha$ vyjadrenej v percentách, obr. 5.4. V našom prípade pre 95%-ný interval spoľahlivosti pre strednú hodnotu je $\Delta = \text{CONFIDENCE.T}(0,05; D9; D17) = 4,491471$.



Obr. 5.4 Dialógové okno analytického nástroja Popisná štatistika so zadáním konfidenčného intervalu a jeho výstup

$P(17,06667 - 4,491471 < M < 17,06667 + 4,491471) = 1 - 0,05$, t.j. s 95%-nou spoľahlivosťou sa stred rozdelenia ZS nachádza v intervale (12,5752; 21,55814).

Pre intervalový odhad rozptylu σ^2 ZS platí $P\left(\frac{(n-1)s_{n-1}^2}{\chi_{n-1}^{2,kv}\left(1-\frac{\alpha}{2}\right)} < \sigma^2 < \frac{(n-1)s_{n-1}^2}{\chi_{n-1}^{2,kv}\left(\frac{\alpha}{2}\right)}\right) = 1 - \alpha$.

Hodnoty chí-kvadrát kvantilov sú $\chi_{n-1}^{2,kv}\left(1 - \frac{\alpha}{2}\right) = \text{CHISQ.INV}(0,975; 14) = 26,11895$ a

$\chi_{n-1}^{2,kv}\left(\frac{\alpha}{2}\right) = \text{CHISQ.INV}(0,025; 14) = 5,628726$. Ich výpočet v Exceli môžeme vidieť na nasledujúcom obrázku 5.5.

	A	B	C	D	E
4	12				
5	12	Mean		17,0666667	
6	12	Standard Error		2,09413391	
7	20	Median		21	
8	21	Mode		12	
9	22	Standard Deviation		8,11054575	
10	22	Sample Variance		65,7809524	
11	23	Kurtosis		-1,0271244	
12	23	Skewness		-0,6900875	
13	25	Range		24	
14	25	Minimum		2	
15	26	Maximum		26	
16		Sum		256	
17		Count		15	
18		Confidence Level(95,0%		4,49147053	
19					
20					
21			chi kv(0,975;14)=	=CHISQ.INV(0,975;14)	
22			chi kv(0,025;14)=	=CHISQ.INV(0,025;14)	
23			95% is pre sigma^2		
24			DH=	=(D17-1)*D10/D21	
25			HH=	=(D17-1)*D10/D22	

A	B	C	D
12			
12	Mean		17,06667
12	Standard Error		2,094134
20	Median		21
21	Mode		12
22	Standard Deviation		8,110546
22	Sample Variance		65,78095
23	Kurtosis		-1,02712
23	Skewness		-0,69009
25	Range		24
25	Minimum		2
26	Maximum		26
	Sum		256
	Count		15
	Confidence Level(95,0%		4,491471
		chi kv(0,975;14)=	26,11895
		chi kv(0,025;14)=	5,628726
		95% is pre sigma^2	
		DH=	35,2592
		HH=	163,6131

Obr. 5.5 Výpočet intervalu spoľahlivosti pre σ^2 v Exceli

$P(35,2592 < \sigma^2 < 163,6131) = 1 - 0,05$, t.j. s 95%-nou spoľahlivosťou sa rozptyl ZS nachádza v intervale (35,2592; 163,6131).

Pre intervalový odhad smerodajnej odchýlky σ ZS platí:

$$P\left(\sqrt{\frac{(n-1)s_{n-1}^2}{\chi_{n-1}^{2,kv}\left(1-\frac{\alpha}{2}\right)}} < \sigma < \sqrt{\frac{(n-1)s_{n-1}^2}{\chi_{n-1}^{2,kv}\left(\frac{\alpha}{2}\right)}}\right) = 1 - \alpha, \quad \text{t.j.} \quad P(\sqrt{35,2592} < \sigma < \sqrt{163,6131}) =$$

0,95. S 95%-nou spoľahlivosťou sa smerodajná odchýlka ZS nachádza v intervale (5,937946; 12,79113).

Príklad 5.2 Firma realizovala na vzorke zamestnancov prieskum dĺžky ich hovorov: 3, 0, 1, 4, 2, 3, 1, 0, 4, 2, 3, 5, 1, 0, 1, 3, 4, 5, 1, 2. Určte charakteristiky súboru náhodne vybraných zamestnancov, urobte odhady charakteristík všetkých zamestnancov firmy. Aká je priemerná dĺžka hovorov zamestnancov firmy? Porovnajte 95%-ný a 90%-ný interval spoľahlivosti pre dĺžku hovorov všetkých zamestnancov.

Riešenie.

Na riešenie príkladu použijeme Excel. Spôsob výpočtu s použitím analytického nástroja Popisná štatistika (Descriptive Statistics) a vybraných funkcií je naznačený na nasledujúcom obr. 5.6.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	3	0		Priklad.	Firma realizovala prieskum dlzky hovorov jej zamestnancov.										
2	0	0													
3	1	0									opisna statistika cez analyticke nástroje				
4	4	1		charakteristiky VS			bodove odhady ZS				Column1				
5	2	1		n	=K18		est M	=K6			Mean	2,25			
6	3	1		xmax	=K16		est σ^2	=K11	=VAR.S(A:A)		Standard Error	0,36			
7	1	1		xmin	=K15		est σ	=K10	=STDEV.S(A:A)		Median	2			
8	0	1		R	=K14						Mode	1			
9	4	2		xpriem	=K6						Standard Deviat	1,62			
10	2	2		s^2	=DEVSQ(A:A)/20	=VAR.P(A:A)					Sample Variance	2,62			
11	3	2		s	=SQRT(E10)	=STDEV.P(A:A)					Kurtosis	-1,09			
12	5	3		k. spicatos	=K12						Skewness	0,21			
13	1	3		k. sikmosti	=K13						Range	5			
14	0	3		modus	=K9						Minimum	0			
15	1	3		median	=K8						Maximum	5			
16	3	4		dolny kvart	=QUARTILE.INC(A:A;1)						Sum	45			
17	4	4		horny kvart	=QUARTILE.INC(A:A;3)						Count	20			
18	5	4													
19	1	5		intervalove odhady ZS pre strednu hodnotu										Column1	
20	2	5		95% IS pre dlzku vsetkych hovorov											
21							delta=	=CONFIDENCE.T(0,05;K10;K18)					Conf	0,75732	
22							95% IS pre M=	=K6-F21	=K6+F21	s 95% pravdepodobnostou dlzka vsetkych hovorov bude z tohoto intervalu					
23													Column1		
24															
25				90% IS pre dlzku vsetkych hovorov											
26							delta=	=CONFIDENCE.T(0,1;K10;20)					Conf	0,62565	
27							90% IS pre M=	=K6-F26	=K6+F26	s 90% pravdepodobnostou dlzka vsetkych hovorov bude z tohoto intervalu					
28															
29				cim je spolahlivost odhadu vyssia, tym je IS sirsi + cim je rozsah VS vacsi, tym je IS uzsi											

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	3	0		Priklad.	Firma realizovala prieskum dlzky hovorov jej zamestnancov.										
2	0	0													
3	1	0									opisna statistika cez analyticke nástroje				
4	4	1		charakteristiky VS			bodove odhady ZS				Column1				
5	2	1		n	20		est M	2,25			Mean	2,25			
6	3	1		xmax	5		est σ^2	2,61842	2,61842		Standard Error	0,36			
7	1	1		xmin	0		est σ	1,61815	1,61815		Median	2			
8	0	1		R	5						Mode	1			
9	4	2		xpriem	2,25						Standard Deviat	1,62			
10	2	2		s^2	2,488	2,4875					Sample Variance	2,62			
11	3	2		s	1,577	1,57718					Kurtosis	-1,09			
12	5	3		k. spicatos	-1,09						Skewness	0,21			
13	1	3		k. sikmosti	0,212						Range	5			
14	0	3		modus	1						Minimum	0			
15	1	3		median	2						Maximum	5			
16	3	4		dolny kvart	1						Sum	45			
17	4	4		horny kvart	3,25						Count	20			
18	5	4													
19	1	5		intervalove odhady ZS pre strednu hodnotu										Column1	
20	2	5		95% IS pre dlzku vsetkych hovorov											
21							delta=	0,75732					Conf	0,75732	
22							95% IS pre M=	1,49268	3,00732	s 95% pravdepodobnostou dlzka vsetkych hovorov bude z tohoto intervalu					
23													Column1		
24															
25				90% IS pre dlzku vsetkych hovorov											
26							delta=	0,62565					Conf	0,62565	
27							90% IS pre M=	1,62435	2,87565	s 90% pravdepodobnostou dlzka vsetkych hovorov bude z tohoto intervalu					
28															
29				cim je spolahlivost odhadu vyssia, tym je IS sirsi + cim je rozsah VS vacsi, tym je IS uzsi											

Obr. 5.6 Výpočet realizovaný v Exceli

Z porovnání 95%-ného a 90%-ného intervalu spoľahlivosti pre dĺžku hovorov všetkých zamestnancov vidíme, že čím je spoľahlivosť odhadu vyššia, tým je interval spoľahlivosti širší.

Príklad 5.3 Pri skúškach životnosti 500 kusov náhodne vybraných pneumatík určitého výrobcu sme získali tieto výsledky v tisícoch kilometrov:

30,0	30,1	36,5	39,8	53,4	39,7	61,8	70,0	52,4	33,4	31,8	38,5	65,3	47,6	47,1	52,0	47,2	66,4	70,0	46,3
45,6	40,0	37,9	45,6	70,2	72,5	59,8	50,3	60,1	63,3	50,3	47,1	52,4	62,4	63,4	49,6	46,6	53,6	56,6	68,2
36,8	37,9	40,2	40,0	52,4	59,8	56,6	59,7	60,1	62,5	50,3	52,4	50,3	51,3	54,3	52,1	58,1	68,0	52,4	30,5
52,4	39,2	38,3	37,9	47,1	37,9	45,2	51,6	50,8	51,8	40,2	50,5	40,2	50,8	50,3	61,8	63,3	71,2	59,8	56,6
66,5	45,2	37,9	72,5	65,4	56,6	70,2	47,1	63,4	51,2	51,3	52,4	30,5	52,4	47,1	51,8	50,8	70,2	72,5	59,8
37,9	51,6	40,2	59,8	47,1	56,7	72,5	59,8	40,0	70,0	58,6	53,4	65,3	62,5	52,4	62,5	50,3	62,5	50,0	50,1
40,2	50,8	40,0	40,0	30,8	52,4	62,5	52,4	50,3	47,1	52,4	51,8	65,3	50,3	47,1	52,4	61,8	50,3	62,5	65,4
40,0	51,4	50,4	70,0	51,2	55,3	47,1	39,2	53,6	55,3	25,3	50,8	47,1	62,5	54,3	55,4	61,3	50,3	53,6	53,6
53,1	51,2	55,8	58,6	53,6	51,8	37,9	38,5	57,3	33,8	36,0	56,1	65,0	50,0	53,2	53,1	63,3	71,2	70,2	37,9
39,8	45,6	40,0	37,9	72,5	39,8	41,2	37,9	56,3	54,1	34,0	39,8	45,6	40,0	37,9	72,5	59,8	40,0	70,0	56,6
30,0	45,6	36,8	52,4	38,5	37,9	40,2	40,0	53,1	39,8	38,5	37,9	40,2	38,5	37,9	40,2	40,0	50,4	53,8	40,0
56,1	39,8	51,4	70,2	52,4	47,1	63,4	47,1	50,8	51,2	51,6	72,5	38,5	62,5	50,3	62,3	50,0	40,0	38,5	50,4
53,4	70,2	52,4	47,1	63,4	47,1	50,8	51,2	53,8	72,5	38,5	52,4	65,3	52,4	50,3	40,2	30,5	65,3	65,3	47,1
47,1	63,4	54,3	50,3	47,1	52,4	47,1	54,1	55,2	37,9	50,3	50,3	52,4	47,1	54,1	55,2	37,9	37,9	50,3	
39,8	45,6	40,0	37,9	70,2	59,8	40,0	70,0	56,6	37,9	52,4	56,2	47,1	50,3	40,0	50,0	30,5	55,2	30,0	35,5
40,0	37,9	39,2	45,2	51,6	50,8	51,4	51,2	45,6	45,8	39,8	70,2	63,4	45,6	47,1	37,9	42,2	40,2	63,4	50,0
59,0	58,6	50,1	45,8	40,0	37,9	72,5	59,8	40,0	70,0	58,6	37,9	52,4	70,2	47,1	50,3	37,9	45,2	50,0	49,5
50,3	59,7	51,8	47,1	59,8	52,4	39,2	38,5	37,9	40,0	47,1	51,2	54,1	57,7	51,2	39,8	30,5	41,2	37,9	46,7
53,6	37,9	56,6	40,0	50,4	47,1	50,3	35,5	50,0	49,5	46,7	35,5	36,0	54,0	38,5	40,2	54,1	50,0	37,9	50,0
40,0	50,0	30,5	55,2	30,5	41,2	37,9	30,5	54,1	54,1	39,8	51,2	72,5	37,9	37,9	45,6	56,3	40,0	49,5	50,2
55,5	64,2	35,8	35,7	35,8	55,5	41,2	40,2	63,4	50,8	47,1	40,0	51,4	41,2	40,2	39,2	37,9	47,1	47,1	37,5
38,5	37,9	39,2	51,2	35,7	64,2	55,6	57,3	58,3	51,1	50,8	51,6	55,2	56,6	45,6	40,0	37,9	50,0	54,1	35,2
46,6	47,0	49,6	49,3	49,9	50,4	52,5	50,0	51,2	52,3	51,3	51,0	52,0	42,0	30,0	54,0	52,1	51,2	54,3	54,2
40,0	50,4	47,1	50,3	35,5	50,0	49,5	46,7	54,1	36,0	51,2	38,5	37,9	40,2	45,6	30,0	40,0	47,1	50,0	51,2
40,0	53,1	39,8	30,0	56,1	53,4	47,1	39,8	40,0	59,0	50,3	51,6	40,0	55,5	38,5	48,6	40,0	30,0	45,6	71,8

Pomocou programu Excel realizujte výpočet základných štatistických charakteristík daného výberového súboru.

Riešenie. Pri výpočte základných charakteristík jedného výberového súboru Excel vyžaduje, aby boli vstupné údaje zapísané v jednom stĺpci, resp. jednom riadku. Výstupom analytického nástroja Popisná štatistika (Descriptive statistics) je tabuľka na obr. 5.7:

	A	B	C	D	E
1	30,0				
2	45,6		Column1		
3	36,8				
4	52,4		Mean	49,7618	
5	38,5		Standard Error	0,448164	
6	37,9		Median	50,3	
7	40,2		Mode	37,9	
8	40,0		Standard Deviation	10,02125	
9	53,1		Sample Variance	100,4255	
10	39,8		Kurtosis	-0,33291	
11	30,0		Skewness	0,310975	
12	56,1		Range	50	
13	53,4		Minimum	30	
14	47,1		Maximum	80	
15	39,8		Sum	24880,9	
16	40,0		Count	500	
17	59,0		Largest(1)	80	
18	50,3		Smallest(1)	30	
19	53,6		Confidence Level(95	0,880521	
20	40,0				

Obr. 5.7 Výstup analytického nástroja Popisná štatistika

Interpretácia výslednej tabuľky je nasledovná:

Bodový odhad strednej hodnoty (Mean) ZS je $estM = \bar{x} = 49\,761,8$ kilometrov – je to priemerná životnosť skúmaných, ale aj všetkých pneumatík daného výrobcu – ak by všetky pneumatiky mali rovnakú životnosť, bola by rovná 49 761,8 kilometrov.

Štandardná chyba (Standard Error) je 0,448164. Štandardná chyba je daná podielom smerodajnej odchýlky a odmocniny rozsahu súboru ($0,448164 = D8/SQRT(D16) = \frac{10,02125}{\sqrt{500}}$).

Medián (Median) je 50 300 kilometrov. Hodnota mediánu delí usporiadaný súbor na dve rovnako početné skupiny, t.j. polovica skúmaných pneumatík mala životnosť maximálne 50 300 tisíc kilometrov a druhá polovica skúmaných pneumatík mala životnosť minimálne 50 300 tisíc kilometrov.

Modus (Mode) je 37 900 kilometrov – najčastejšie sa vyskytujúca dĺžka životnosti pneumatík v skúmanom súbore (najviac pneumatík malo životnosť 37 900 kilometrov).

Poznámka 5.2 Excel uvádza iba jednu hodnotu modusu, a to aj v prípade viacvrcholového rozdelenia početností.

Výberová smerodajná odchýlka (Standard Deviation) 10 021,25 kilometrov udáva priemerné odchýlky od priemernej životnosti pneumatiky 49 761,8 kilometrov.

Výberový rozptyl je $s_n^2 = 100\,425,5 \text{ km}^2$, vyjadruje priemerný štvorec odchýlok od priemernej životnosti pneumatiky 49 761,8 kilometrov.

Koeficient špicatosti (Kurtosis) je -0,33291. To, že jeho hodnota je menšia ako 0, znamená, že rozdelenie početností skúmaného štatistického znaku životnosti pneumatík daného výrobcu je „plochejšie“, ako má normované normálne rozdelenie.

Hodnota koeficientu šikmosti (Skewness) 0,310975 je väčšia ako 0, preto je toto rozdelenie zošikmené doľava.

Variačné rozpätie (Range) s hodnotou 50 tisíc kilometrov je rozdielom maximálnej hodnoty (Maximum) 80 tisíc kilometrov a minimálnej hodnoty (Minimum) 30 -tisíc kilometrov.

Súčet všetkých hodnôt (Sum) je 24880,9 tisíc kilometrov – všetky pneumatiky z výberového súboru prešli spolu v súčte 24880,9 tisíc kilometrov.

Rozsah súboru (Count) je 500 – čo je počet pneumatík, ktoré tvorili výberovú vzorku (výberový súbor).

Prípustná chyba odchýlky (Confidence Level) je 0,880521, t.j. pri zovšeobecnení výsledkov na celý základný súbor, t.j. na všetky pneumatiky skúmaného výrobcu s 95%-nou spoľahlivosťou sa životnosť pneumatík nachádza v intervale 49 761,8 kilometrov $\pm$ 880,521 kilometrov. Môžeme tvrdiť, že s 95%-nou spoľahlivosťou je životnosť pneumatík daného výrobcu od 48 837, 479 kilometrov do 50 642,321 kilometrov.

Porovnanie výstupu analytického nástroja Popisná štatistika s výstupmi jednotlivých štatistických funkcií môžeme vidieť na nasledujúcom obrázku 5.8:

	A	B	C	D	E	F	G	H
1	30,0							
2	45,6		Column1					
3	36,8							
4	52,4	Mean		49,7618	=AVERAGE(A1:A500)			
5	38,5	Standard Error		0,448164	=D8/SQRT(D16)	=E8/E16^(1/2)		
6	37,9	Median		50,3	=MEDIAN(A1:A500)			
7	40,2	Mode		37,9	=MODE.MULT(A1:A500)			
8	40,0	Standard Deviation		10,02125	=STDEV.S(A1:A500)			
9	53,1	Sample Variance		100,4255	=VARA(A1:A500)			
10	39,8	Kurtosis		-0,33291	=KURT(A1:A500)			
11	30,0	Skewness		0,310975	=SKEW(A1:A500)			
12	56,1	Range		50	=MAX(A1:A500)-MIN(A1:A500)			
13	53,4	Minimum		30	=MIN(A1:A500)			
14	47,1	Maximum		80	=MAX(A1:A500)			
15	39,8	Sum		24880,9	=SUM(A1:A500)			
16	40,0	Count		500	=COUNT(A1:A500)	=FREQUENCY(A1:A500;80)		
17	59,0	Largest(1)		80				
18	50,3	Smallest(1)		30				
19	53,6	Confidence Level(95		0,880521	=CONFIDENCE.NORM(0,05;E8;E16)	=CONFIDENCE.T(0,05;E8;E16)		

	A	B	C	D	E	F
1	30,0					
2	45,6		Column1			
3	36,8					
4	52,4	Mean		49,7618	49,8	
5	38,5	Standard Error		0,448164	0,448164	0,448164
6	37,9	Median		50,3	50,3	
7	40,2	Mode		37,9	37,9	
8	40,0	Standard Deviation		10,02125	10,02125	
9	53,1	Sample Variance		100,4255	100,4255	
10	39,8	Kurtosis		-0,33291	-0,33291	
11	30,0	Skewness		0,310975	0,310975	
12	56,1	Range		50	50,0	
13	53,4	Minimum		30	30,0	
14	47,1	Maximum		80	80,0	
15	39,8	Sum		24880,9	24880,9	
16	40,0	Count		500	500	500
17	59,0	Largest(1)		80		
18	50,3	Smallest(1)		30		
19	53,6	Confidence Level(95		0,880521	0,878385	0,880521

Obr. 5.8 Porovnanie výstupu analytického nástroja Popisná štatistika s výstupmi jednotlivých štatistických funkcií

Úlohy

5.1 Nech výberový súbor je: 3, 5, 8, 9, 11, 13, 15, 17, 19, 21, 22, 22, 23, 25. Nájdite opisné číselné charakteristiky výberového súboru a 95 %-ný interval spoľahlivosti pre strednú hodnotu ZS.

$[n = 14; x_{\max} = 25; x_{\min} = 3; R = 22; \bar{x} = 15,214; s_n^2 = 47,597; s_n = 6,899; \text{k.špicatosti} = -1,223; \text{k.šikmosti} = -0,318; \text{modus} = 22; \text{medián} = 15,214; P(11,081 < M < 19,348) = 0,95]$

5.2 V 16 vzorkách ocele sa zisťoval obsah kobaltu v % s výsledkami: 1,52; 1,92; 2,75; 2,21; 1,84; 2,15; 2,36; 2,59; 1,89; 2,11; 2,23; 2,62; 2,30; 2,52; 1,47; 2,52. Vypočítajte priemerný obsah kobaltu v % v skúmaných 16 vzorkách, vypočítajte rozptyl a smerodajnú odchýlku VS.

$[\bar{x} = 2,19; s_n^2 = 0,14; s_n = \sqrt{0,14}]$

5.3 Náhodným výberom bola vybraná vzorka výrobkov, na ktorých sa skúmala hmotnosť v gramoch: 13, 11, 19, 11, 22, 22, 13, 11, 17, 13, 27, 14, 16, 13, 24, 30, 9, 12, 15, 15, 21, 18, 11, 9, 13. Určte 95%-ný interval spoľahlivosti pre priemernú hmotnosť všetkých výrobkov. V akom intervale sa bude nachádzať 90% hmotnosti všetkých výrobkov?

[(13,40174 g; 17,99826 g); (13,79484 g; 17; 60516 g)]

5.4 Na úseku cesty s doporučenou rýchlosťou 80 km/h sme kontrolovali rýchlosť náhodne vybraných motorových vozidiel. Získali sme údaje: 72, 73, 65, 136, 72, 73, 66, 73, 73, 72. Pri predpoklade, že rozdelenie rýchlostí idúcich vozidiel možno považovať za normálne, určte bodový a 95%-ný intervalový odhad rýchlosti automobilov.

[77,5 km/h; (62,6 km/h; 92,4 km/h)]

5.2 TRIEDENIE

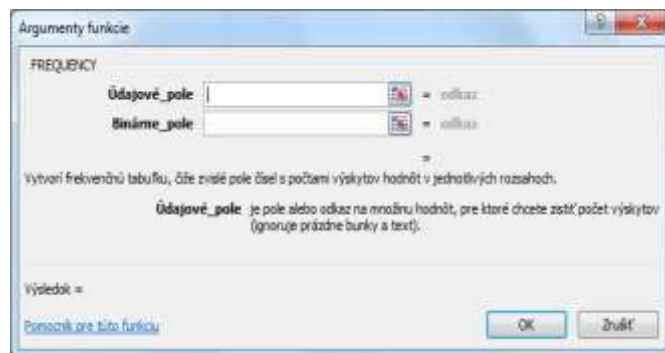
5.2.1 Jednoduché triedenie

Pri analýze výberového súboru, t.j. pri analýze nameraných hodnôt štatistického znaku, ktoré získame skúmaním vlastností náhodne vybraných štatistických jednotiek zo základného súboru, určujeme rad rozdelenia početností výskytu hodnôt skúmanej premennej, ktorý je empirickým ekvivalentom teoretického zákona rozdelenia pravdepodobnosti. Získame ho triedením. Rad rozdelenia početností výskytu hodnôt skúmanej premennej špecifikuje, ako často, s akou frekvenciou sa vyskytujú jednotlivé hodnoty, resp. triedy hodnôt.

Príklad 5.4 Počty výrobkov vyrobených jednotlivými vybranými pracovníkmi za zmenu sú: 5, 1, 5, 2, 4, 5, 4, 2, 5, 1, 4, 4, 5, 5, 2, 5, 4, 9, 4, 5. Zostrojte frekvenčnú tabuľku a zobrazte histogram rozdelenia početností počtu vyrobených výrobkov.

Riešenie. Rad rozdelenia početností počtu výrobkov, vyrobených jednotlivými pracovníkmi zapíšeme pomocou frekvenčnej tabuľky, ktorá bude obsahovať nasledujúce údaje: x_i označuje hodnotu štatistického znaku v i -tej triede, n_i sú absolútne početnosti, N_i kumulatívne absolútne početnosti, f_i relatívne početnosti, F_i sú kumulatívne relatívne početnosti.

Štatistický znak - počet výrobkov, vyrobených jednotlivými pracovníkmi za zmenu, nadobúda „malý“ počet diskrétnych hodnôt, preto môžeme použiť jednoduché triedenie. Celý súbor rozdelíme na päť tried, pričom každú triedu bude reprezentovať iná hodnota štatistického znaku x_i . Absolútne početnosti n_i určujú počet štatistických jednotiek v danej i -tej triede, t.j. pri jednoduchom triedení počet výskytov určitej hodnoty štatistického znaku v súbore nameraných hodnôt. Početnosti výskytov určitého znaku určíme pomocou triedenia, pričom dodržiavame zásady jednoznačnosti (každá hodnota musí byť zaradená iba do jednej triedy) a úplnosti (každá hodnota musí byť zaradená do niektorej z tried). V Exceli môžeme na to použiť preddefinovanú funkciu `=FREQUENCY(údajové_pole;binárne_pole)`, kde do položky `údajové_pole` zadávame všetky namerané hodnoty štatistického znaku a do položky `binárne_pole` zadávame jednotlivé hodnoty štatistického znaku reprezentujúce danú triedu. Jej dialógové okno je na nasledujúcom obr. 5.9.



Obr. 5.9 Dialógové okno funkcie FREQUENCY

Použitie tejto funkcie je naznačené v nasledujúcej časti. Pred samotným triedením je potrebné vyznačiť bunky, do ktorých chceme výsledky triedenia zapísať. Funkcia FREQUENCY je definovaná tak, že prvú triedu reprezentujú hodnoty menšie alebo rovné prvej hodnote (prvému číslu) v binárnom poli, druhú triedu reprezentujú hodnoty od prvej po druhú hodnotu v binárnom poli, ..., predposlednú triedu hodnoty od predposledného po posledné číslo v binárnom poli. Ak označíme výsledné pole, väčšie ako binárne pole, poslednou triedou sú hodnoty väčšie ako posledné číslo tohto binárneho poľa (slúži na kontrolu). Funkciu odosielame súčasným stlačením kláves CTRL, SHIFT, ENTER. Pri jednoduchom triedení takto získame všetky početnosti jednotlivých hodnôt štatistického znaku. Ak by sme funkciu potvrdili iba klávesom ENTER, resp. OK v dialógovom okne, potom funkcia FREQUENCY vypočíta iba absolútnu triednu početnosť pre prvú triedu. Použitie funkcie a jej výstup je zobrazený na nasledujúcom obr. 5.10.

až po označení výstupnej oblasti voláme funkciu FREQUENCY

	A	B	C	D	E	F	G	H
1	5		Příklad.	Počet výrobkov, vyrobených jednotlivými pracovníkmi				
2	1							
3	5		x_i	n_i	N_i	f_i	F_i	
4	2		1					
5	4		2					
6	5		4					
7	4		5					
8	2		9					
9	5							
10	1							
11	4							
12	4							
13	5							
14	5							
15	2							
16	5							
17	4							
18	9							
19	4							
20	5							

ako prvé označíme bunky, do ktorých chceme zapísať početnosti jednotlivých tried

	x_i	n_i	N_i	f_i	F_i
1					
2					
4					
5					
9					

$=FREQUENCY$

$FREQUENCY(Údajové_pole; binárne_pole)$

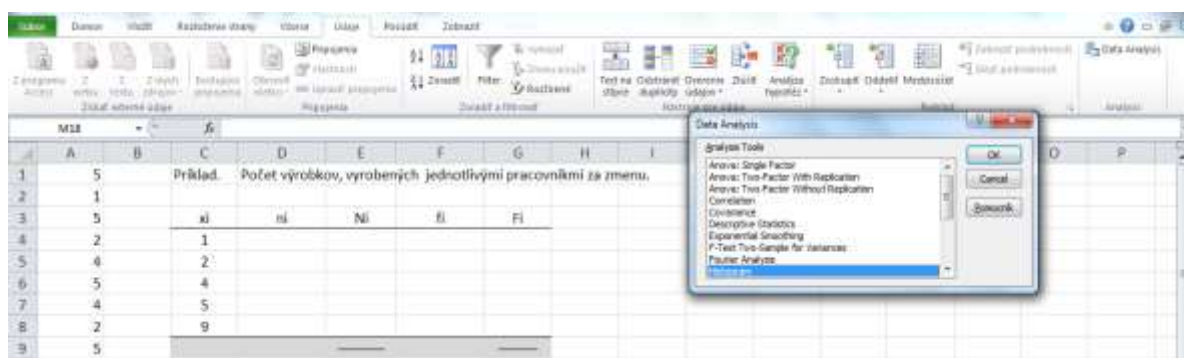
príkaz potvrdíme trojkombináciou kláves CTRL +SHIFT+ENTER

	x_i	n_i	N_i	f_i	F_i
1		2			
2		3			
4		6			
5		8			
9		1			

výstupom sú vypočítané triedne početnosti

Obr. 5.10 Triedenie pomocou funkcie FREQUENCY

Rozdelenia početností môžeme získať aj pomocou analytického nástroja Histogram:



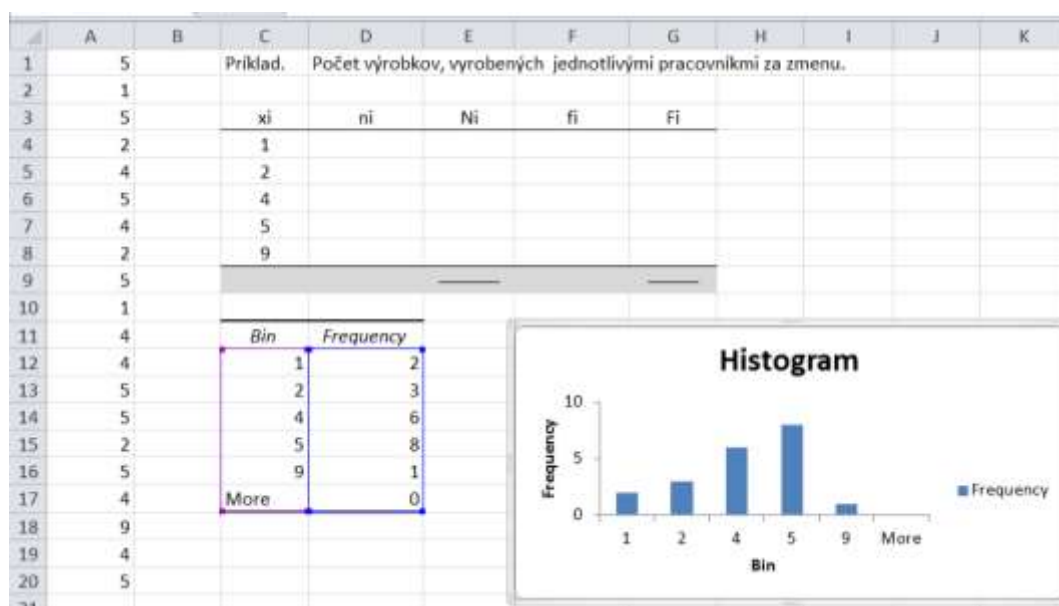
Obr. 5.11 Triedenie pomocou Analýzy dát

Po voľbe Údaje→Analýza dát→Histogramu (Data→Data analysis→Histogram) sa objaví dialógové okno tohto analytického nástroja:



Obr. 5.12 Dialógové okno analytického nástroja Histogram so zadani vstupných údajov

Princíp triedenia je rovnaký ako pri funkcii FREQUENCY. Dostaneme výstup, ktorého súčasťou je okrem radu rozdelenia početností aj jeho grafické znázornenie – histogram :



Obr. 5.13 Triedenie pomocou analytického nástroja Histogram

Rad rozdelenia početností počtu vyrobených výrobkov zobrazíme v tvare frekvenčnej tabuľky (obr. 5.14), kde triedy x_i sú počty výrobkov, vyrobených pracovníkmi počas zmeny, absolútne početnosti n_i špecifikujú, koľko statistických jednotiek (zamestnancov) prislúcha danej triede, t.j. dvaja zamestnanci vyrobili počas zmeny 1 výrobok, traja zamestnanci vyrobili počas zmeny 2 výrobky, šiesti zamestnanci 4 výrobky...

Kumulatívne absolútne početnosti N_i reprezentujú početnosť výskytu hodnôt menších alebo rovných príslušnej triede, t.j. dvaja zamestnanci vyrobili počas zmeny 1 výrobok, piati zamestnanci vyrobili 1 alebo 2 výrobky (maximálne 2 výrobky), jedenásti zamestnanci vyrobili maximálne 6 výrobkov...

Relatívne početnosti f_i vyjadrujú podiel (pravdepodobnosť) výskytu hodnôt príslušnej triedy na celkovom počte analyzovaných hodnôt, to znamená 10% zamestnancov vyrobilo počas zmeny 1 výrobok, 15% zamestnancov vyrobilo 2 výrobky...

Kumulatívne relatívne početnosti F_i zodpovedajú distribučnej funkcii z teórie pravdepodobnosti, čítame 10% zamestnancov vyrobilo počas zmeny 1 výrobok, 25% zamestnancov vyrobilo počas zmeny maximálne 2 výrobky, ...

Posledný súčtový riadok frekvenčnej tabuľky slúži na kontrolu správnosti riešenia.

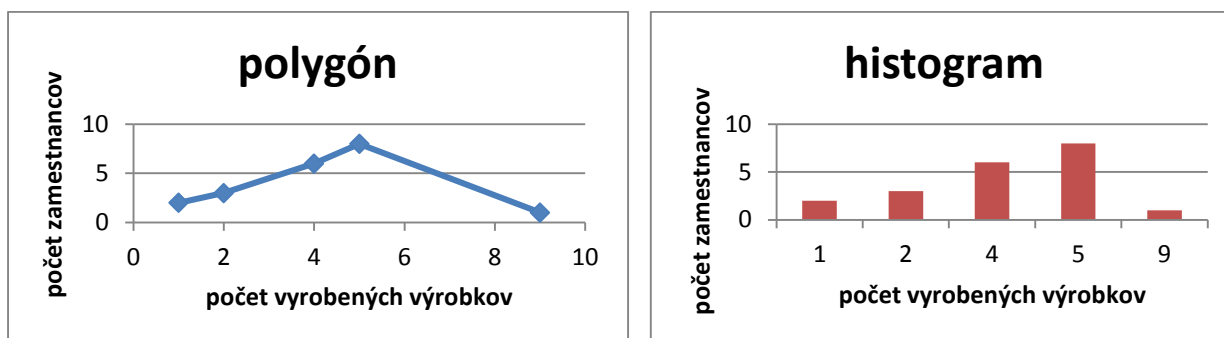
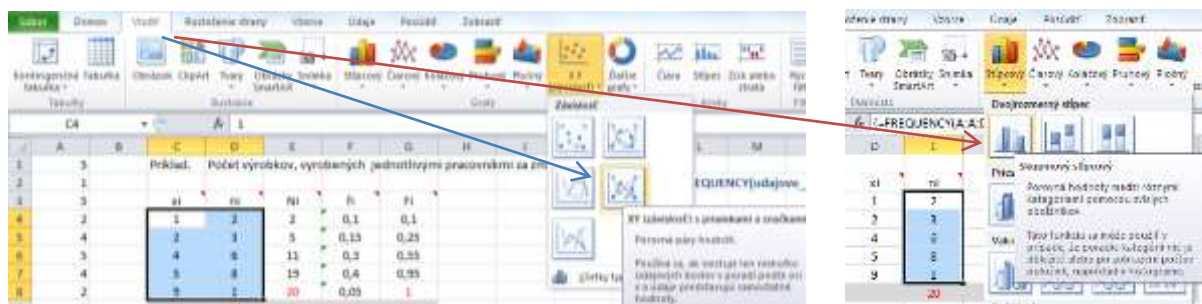
	A	B	C	D	E	F	G
1	5		Příklad.	Počet výrobkov, vyrobených jednotlivými pracovníkmi za zmenu.			
2	1						
3	5		x_i	n_i	N_i	f_i	F_i
4	2		1	=FREQUENCY(A1:A20;C4:C8)	=D4	=D4/\$D\$9	=F4
5	4		2	=FREQUENCY(A1:A20;C4:C8)	=E4+D5	=D5/\$D\$9	=G4+F5
6	5		4	=FREQUENCY(A1:A20;C4:C8)	=E5+D6	=D6/\$D\$9	=G5+F6
7	4		5	=FREQUENCY(A1:A20;C4:C8)	=E6+D7	=D7/\$D\$9	=G6+F7
8	2		9	=FREQUENCY(A1:A20;C4:C8)	=E7+D8	=D8/\$D\$9	=G7+F8
9	5			=SUM(D4:D8)		=SUM(F4:F8)	

3	5		x_i	n_i	N_i	f_i	F_i
4	2		1	2	2	0,1	0,1
5	4		2	3	5	0,15	0,25
6	5		4	6	11	0,3	0,55
7	4		5	8	19	0,4	0,95
8	2		9	1	20	0,05	1
9	5		Σ :	20		1	

Obr. 5.14 Výsledná frekvenčná tabuľka so zobrazením použitých funkcií v Exceli

Grafickou reprezentáciou radu rozdelenia početností možných hodnôt – tried možných hodnôt je polygón a histogram. Polygón vytvoríme v Exceli ako spojnicový graf, histogram ako stĺpcový graf. Spôsob ich vytvárania spolu s výstupom sú na nasledujúcom obrázku 5.15.

Po označení zobrazovaných hodnôt volíme postupnosť krokov Vložiť→Grafy→XY závislosť→XY (závislosť) s priamkami a značkami pri zobrazovaní polygónu, resp. Vložiť→Grafy→Stĺpcový→Dvojrozmerný stĺpec→Skupinový stĺpcový pri zobrazovaní histogramu.



Obr. 5.15 Grafická reprezentácia radu rozdelenia početností - polygón a histogram

5.2.2 Intervalové triedenie

Príklad 5.5 Dĺžka vybraných výrobkov v cm je: 97, 103, 104, 106, 106, 111, 111, 112, 112, 112, 113, 113, 114, 114, 114, 115, 117, 118, 118, 118, 118, 119, 121, 121, 124, 124, 125, 125, 126, 127, 127, 129, 129, 129, 132, 132, 133, 133, 136, 151. Roztriedte uvedené údaje do intervalov, za dolnú hranicu prvého intervalu si zvolte 90 cm a šírku intervalov 10 cm. Zostrojte frekvenčnú tabuľku, histogram a rad rozdelenia početností dĺžky vybraných výrobkov.

Riešenie. Ak náhodná premenná nadobúda veľký počet rôznych hodnôt, volíme triedenie do navzájom disjunktných intervalov. Najbežnejší je spôsob značenia polouzavretého intervalu (DH, HH) , kde DH je dolná hranica intervalu, HH označuje jeho hornú hranicu. Intervaly sú zvyčajne ekvidistantné (rovnakej veľkosti). Existuje niekoľko spôsobov na stanovenie počtu k intervalov, napr. $k \leq 5 \log n$, $k \approx \sqrt{n}$, Sturgessovo pravidlo $k \approx 1 + 3,322 \log n$, Yuleho vzorec $k \approx 2,5 \sqrt[4]{n}$, Prioritou je vždy vychádzať z prirodzeného, objektívnou skutočnosťou predurčeného počtu tried. V Exceli triedime pomocou funkcie `=FREQUENCY(údajové_pole;binárne_pole)`.

Pri intervalovom triedení do položky `binárne_pole` zadávame hodnoty, ktoré reprezentujú horné hranice triednych intervalov. Keďže funkcia triedi podľa už skôr popísaného princípu, je dôležité dodržať toto pravidlo. Pri nesprávnom označení binárneho poľa triedenie do intervalov nebude správne. Pri vyznačovaní oblasti, do ktorej chceme vypísať výsledky triedenia, môžeme vyznačiť navyše jeden riadok, ktorý bude určený pre hodnoty štatistického znaku väčšieho ako posledná hodnota binárneho poľa (služi na kontrolu) alebo výsledok triedenia môžeme

skontrolovať súčtom všetkých absolútnych početností, ktorý musí byť rovný rozsahu VS. Postup triedenia je naznačený na nasledujúcom obr. 5.16.

	A	B	C	D	E	F
1	97		Příklad.	Délka vybraných výrobků v cm.		
2	103					
3	104		DH	HH	ni	
4	106		90	100		
5	106		100	110		
6	111		110	120		
7	111		120	130		
8	112		130	140		
9	112		140	150		
10	112		150	160		
11	113					
12	113					

	A	B	C	D	E	F
1	97		Příklad.	Délka vybraných výrobků v cm.		
2	103					
3	104		DH	HH	ni	
4	106		90	100		1
5	106		100	110		4
6	111		110	120		17
7	111		120	130		12
8	112		130	140		5
9	112		140	150		0
10	112		150	160		1
11	113					0

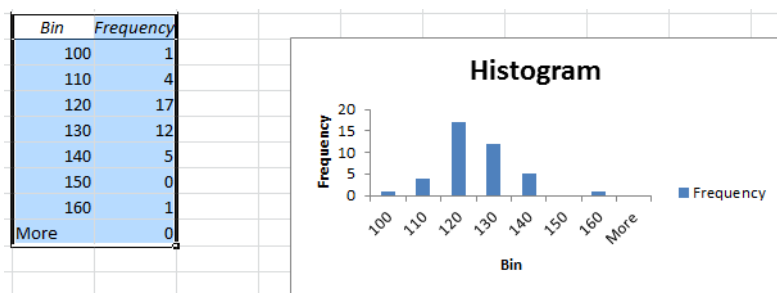
Obr. 5.16 Triedenie pomocou funkcie =FREQUENCY

Druhým spôsobom získania rozdelenia početností je triedenie pomocou analytického nástroja Histogram, ktorý dostaneme postupnosťou príkazov Údaje→Analýza dát→Histogram (Data→Data analysis→Histogram). Jeho dialógové okno je na obrázku 5.17.



Obr. 5.17 Dialógové okno analytického nástroja Histogram so zadáním vstupných údajov

Analytický nástroj Histogram automaticky počíta absolútne početnosti výskytu príslušných tried. Výstup je možné doplniť potvrdením voľby Vytvoriť graf (Chart Output) obrázkom histogramu. Výsledkom intervalového triedenia zadaných hodnôt bude nasledujúci obr. 5.18.



Obr. 5.18 Výstup intervalové triedenia realizovaného pomocou Analýzy dát

Rad rozdelenia početností, reprezentovaný frekvenčnou tabuľkou je zobrazený na obr. 5.19.

	A	B	C	D	E	F	G	H
1	97		Příklad.	Dĺžka vybraných výrobkov v cm.				
2	103							
3	104		DH	HH	ni	Ni	fi	Fi
4	106		90	100	=FREQUENCY(A1:A40;D4:D10)	=E4	=E4/\$E\$11	=G4
5	106		100	110	=FREQUENCY(A1:A40;D4:D10)	=F4+E5	=E5/\$E\$11	=H4+G5
6	111		110	120	=FREQUENCY(A1:A40;D4:D10)	=F5+E6	=E6/\$E\$11	=H5+G6
7	111		120	130	=FREQUENCY(A1:A40;D4:D10)	=F6+E7	=E7/\$E\$11	=H6+G7
8	112		130	140	=FREQUENCY(A1:A40;D4:D10)	=F7+E8	=E8/\$E\$11	=H7+G8
9	112		140	150	=FREQUENCY(A1:A40;D4:D10)	=F8+E9	=E9/\$E\$11	=H8+G9
10	112		150	160	=FREQUENCY(A1:A40;D4:D10)	=F9+E10	=E10/\$E\$11	=H9+G10
11	113		Σ:		=SUM(E4:E10)		=SUM(G4:G10)	

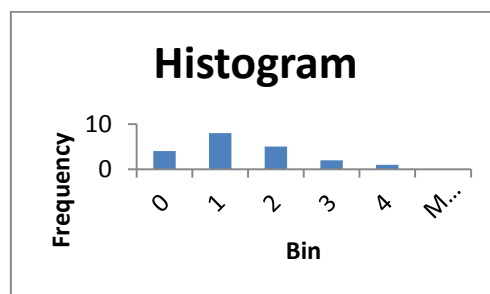
DH	HH	ni	Ni	fi	Fi
90	100	1	1	0,025	0,025
100	110	4	5	0,1	0,125
110	120	17	22	0,425	0,55
120	130	12	34	0,3	0,85
130	140	5	39	0,125	0,975
140	150	0	39	0	0,975
150	160	1	40	0,025	1
Σ:		40		1	

Obr. 5.19 Výsledná frekvenčná tabuľka so zobrazením použitých funkcií v Exceli

Úlohy

5.5 U 20 vybraných pracovníkov sa zistil počet vyrobených nepodarkov za mesiac: 1, 0, 2, 1, 3, 2, 1, 2, 0, 1, 2, 3, 4, 1, 2, 1, 0, 1, 1, 0. Zostrojte frekvenčnú tabuľku a histogram.

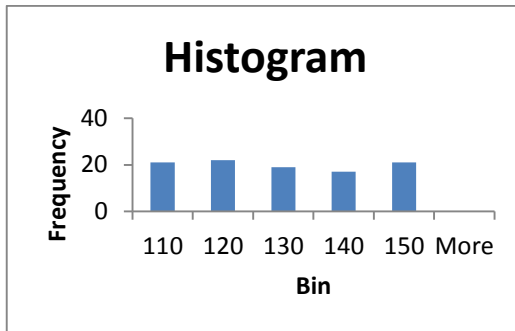
xi	ni	Ni	fi	Fi
0	4	4	0,2	0,2
1	8	12	0,4	0,6
2	5	17	0,25	0,85
3	2	19	0,1	0,95
4	1	20	0,05	1
	20		1	



5.6 Štatistickým znakom je hmotnosť určitého výrobku v gramoch. U vybraných výrobkov boli zistené nasledujúce údaje: 117, 110, 114, 131, 142, 129, 140, 144, 150, 139, 122, 130, 144, 124, 128, 114, 105, 129, 130, 115, 119, 108, 147, 131, 128, 110, 132, 126, 150, 128, 122, 113, 114, 109, 111, 101, 136, 106, 110, 136, 105, 126, 143, 103, 141, 116, 139, 149, 112, 150, 112, 105, 144, 105, 114, 136, 125, 126, 146, 144, 128, 132, 130, 116, 110, 105, 110, 116, 150, 146, 115, 107, 132, 139, 112, 105, 103, 112, 145, 116, 142, 115, 111, 123, 138, 141, 109, 140, 140, 139, 111, 100, 143, 112, 100, 136, 130, 144, 147, 122.

Roztriedte uvedené údaje do intervalov, zostrojte frekvenčnú tabuľku a histogram. Za dolnú hranicu prvého intervalu si zvolte 100 gramov a rozpätie intervalov 10 gramov.

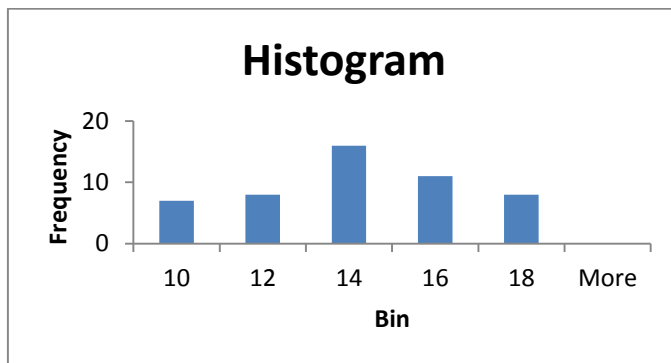
[	DH	HH	ni	Ni	fi	Fi	
	100	110	21	21	0,21	0,21	
	110	120	22	43	0,22	0,43	
	120	130	19	62	0,19	0,62	
	130	140	17	79	0,17	0,79	
	140	150	21	100	0,21	1	
Σ			100		1		]



5.7 Počas 50 týždňov bola sledovaná výroba panelov. Počty nekvalitných panelov počas jednotlivých týždňov sú: 13, 10, 14, 16, 15, 12, 17, 14, 16, 14, 14, 11, 13, 15, 14, 14, 13, 16, 15, 16, 10, 13, 18, 14, 16, 9, 13, 9, 8, 11, 16, 13, 12, 10, 12, 16, 18, 18, 18, 15, 14, 11, 18, 12, 14, 9, 17, 12, 14, 17.

Roztriedte uvedené údaje do intervalov a vypočítajte kumulatívne absolútne početnosti. Za hornú hranicu prvého intervalu si zvolte 10 a rozpätie intervalov 2.

[							
	<i>Bin</i>	<i>Frequency</i>					
	10	7					
	12	8					
	14	16					
	16	11					
	18	8					
	More	0					
	DH	HH	ni	Ni			
	8	10	7	7			
	10	12	8	15			
	12	14	16	31			
	14	16	11	42			
	16	18	8	50			
			50				



5.3 SKÚMANIE ŠTATISTICKEJ ZÁVISLOSTI DVOCH KVANTITÍVNYCH ZNAKOV (PÁROVÁ LINEÁRNA REGRESIA)

Príčinné (kauzálne) závislosti javov:

- jednostranná závislosť – vzťah medzi príčinou a účinkom je jednostranný;
- obojstranná závislosť – príčina a účinok na seba navzájom pôsobia a navzájom sa ovplyvňujú.

Zdanlivá závislosť javov – vzťah medzi javmi nie je dôsledkom ich vzájomnej príčinnej súvislosti, ale pôsobenia ďalšieho javu.

Ak skúmame závislosti medzi kvantitatívnymi znakmi, hovoríme o korelačnom počte; pri skúmaní závislostí medzi kvalitatívnymi znakmi hovoríme o meraní ich asociácie.

Rozlišujem dve základné úlohy korelačného počtu:

- korelačná analýza – úlohou je posúdiť tesnosť štatistickej závislosti;
- regresná analýza – chceme charakterizovať priebeh závislosti, t.j. odhadnúť funkčný vzťah, podľa ktorého sa mení závislá premenná pri zmenách nezávislej premennej.

Podľa toho, koľko činiteľov (nezávislých premenných) berieme do úvahy rozlišujeme:

- jednoduchú (párovú) koreláciu a regresiu – iba jedna nezávislá premenná;
- viacnásobnú (mnohonásobnú) koreláciu a regresiu – viac ako jedna nezávislá premenná.

Korelačná párová analýza

Vzťah medzi dvoma náhodnými premennými môže mať rôznu intenzitu, od úplnej nezávislosti až po úplnú funkčnú nezávislosť. Stupeň lineárnej štatistickej závislosti budeme popisovať kovarianciou a korelačným koeficientom.

Definícia 5.6 Kovariancia medzi premennými X a Y vo výberovom súbore sa definuje pomocou vzťahu

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

kde x_i, y_i sú hodnoty premenných X a Y v i -tom pozorovaní (na i -tej štatistickej jednotke) a $\bar{x}$ a $\bar{y}$ sú aritmetické priemery premenných X a Y .

Pomocou kovariancie sa môžeme presvedčiť, či sú premenné X a Y vo vzájomnej lineárnej závislosti. Platí:

1. $\text{cov}(x, y) = \text{cov}(y, x) \in R$;
2. $\text{cov}(x, y) = 0$ reprezentuje lineárnu nezávislosť premenných X a Y , nehovorí nič o iných typoch závislostí;
3. $\text{cov}(x, y) > 0$, ak medzi premennými X a Y existuje priama lineárna závislosť (hodnoty premennej X a Y narastajú súčasne);

4. $cov(x, y) < 0$, ak medzi premennými X a Y existuje nepriama lineárna závislosť (s nárastom hodnôt premennej X hodnoty premennej Y klesajú).

Nevýhodou kovariancie je, že jej hodnota závisí od mierky, v ktorej sú vyjadrené hodnoty premenných X a Y . Preto vznikla normovaná kovariancia, tzv. korelačný koeficient r_{xy} .

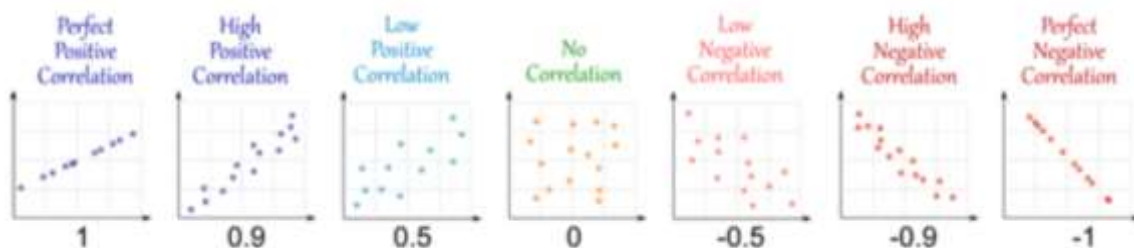
Definícia 5.7 Normovanou mierou vzájomnej lineárnej závislosti – mierou sily, tesnosti lineárneho vzťahu medzi dvoma premennými X a Y je (výberový) koeficient korelácie r_{xy} , ktorý je daný vzťahom $r_{xy} = \frac{cov(x,y)}{s_x \cdot s_y}$.

Vlastnosti korelačného koeficientu sú:

1. $r_{xy} = r_{yx}$; $r_{xy} \in \langle -1, 1 \rangle$;
2. $r_{xy} = 0$ ($cov(x, y) = 0$), ak sú premenné X a Y lineárne nezávislé (nekorelované);
3. $r_{xy} > 0$, ak medzi premennými X a Y existuje priama lineárna závislosť;
4. $r_{xy} < 0$, ak medzi premennými X a Y existuje nepriama lineárna závislosť;
5. ak $|r_{xy}| < 0,3$ hovoríme, že medzi premennými X a Y je malá lineárna závislosť;
6. ak $0,3 \leq |r_{xy}| \leq 0,8$, potom medzi X a Y je mierna (stredná) lineárna závislosť;
7. ak $|r_{xy}| > 0,8$, potom medzi premennými X a Y je silná (tesná) lineárna závislosť;
8. r_{xy} blíži sa k $+1$, znamená vysokú (tesnú) priamu lineárnu závislosť medzi X a Y ;
9. r_{xy} blíži sa k -1 , znamená vysokú (tesnú) nepriamu lineárnu závislosť medzi X a Y ;
10. $|r_{xy}| = 1$, ak medzi premennými X a Y existuje priama lineárna závislosť (namerané hodnoty ležia na modelovej priamke).

Hodnota korelačného koeficientu je nezávislá na merných jednotkách.

Ak medzi náhodnými premennými existuje závislosť, hovoríme, že sú korelované. Korelácia môže byť priama (pozitívna) alebo nepriama (negatívna), obr.5.20.



Obr. 5.20 Zobrazenie korelácie medzi dvoma náhodnými premennými, prevzaté z <https://www.mathsisfun.com/data/correlation.html>

Hodnotu kovariancie vypočítame v Exceli pomocou štatistickej funkcie =COVARIANCE.P (pole1; pole2), hodnotu korelačného koeficientu pomocou funkcie =CORREL (pole1; pole2) alebo funkcie =PEARSON(pole1; pole2), kde pole1 a pole2 predstavujú oblasť 1. premennej a oblasť 2. premennej. Keďže ide o dvojice hodnôt za analyzované premenné, obe oblasti musia mať rovnaký rozmer (rovnakú veľkosť). Na poradií zadávaných oblastí nezáleží.

Regresná analýza

Tvarom (typom) závislosti dvoch náhodných premenných sa zaoberá regresná analýza. Úlohou regresnej analýzy pri skúmaní štatistickej závislosti je nájsť vhodný matematický model (regresnú funkciu), ktorý vyjadruje závislosť medzi závislou premennou Y a nezávislými premennými $X_1, X_2, \dots, X_k$, t.j. zvoliť vhodný typ regresnej funkcie (správne vystihnúť charakter závislosti) a následne odhadnúť parametre regresnej funkcie, pričom východiskom sú empirické (namerané, výberové) údaje. Regresná funkcia môže byť vo všeobecnosti ľubovoľná funkcia $Y = f(\underbrace{X_1, X_2, \dots, X_k}_{\text{nezávislé premenné}}, \underbrace{B_0, B_1, \dots, B_k}_{\text{parametre}})$.

Koeficienty (parametre) $B_0, B_1, \dots, B_p$ odhadneme metódou najmenších štvorcov, t.j. hľadáme najmenšiu hodnotu funkcie $G = G(b_0, b_1, \dots, b_p) = \sum_i e_i^2 = \sum_i (y_i - \hat{y}_i)^2$, kde e_i je náhodná chyba (reziduálna odchýlka, rezíduum), t.j. rozdiel medzi skutočne zistenou (empirickou, nameranou) hodnotou závislej premennej y_i a jej vyrovnanou (teoretickou, odhadovanou, modelovanou) hodnotou $\hat{y}_i = f(x_1, x_2, \dots, x_k, \underbrace{b_0, b_1, \dots, b_k}_{\substack{\text{odhady} \\ \text{parametrov}}})$.

Regresná párová analýza

Ak skúmame na štatistických jednotkách dvojice hodnôt náhodných premenných X a Y , hovoríme o párovej regresii. Regresná funkcia $Y = f(X, B_0, B_1)$. Lineárny vzťah medzi X a Y vyjadríme funkčnou závislosťou $Y = B_0X + B_1$. Bodový odhad označíme vzťahom $\hat{y}_i = b_0 + b_1x_i$. Hľadáme lokálne minimum funkcie $G(b_0, b_1) = \sum_{i=1}^n (y_i - (b_0 + b_1x_i))^2 = \sum_{i=1}^n (y_i - b_0 - b_1x_i)^2$ – t.j. hľadáme minimum súčtu štvorcov odchýlok nameraných a modelovaných hodnôt (minimum súčtu reziduálnych štvorcov). Nepoznáme b_0, b_1 . Nutná podmienka existencie extrémů hovorí, že musí platiť:

$$\begin{aligned}\frac{\partial G}{\partial b_0} &= \sum_{i=1}^n 2(y_i - b_0 - b_1x_i) \cdot (-1) = 0 \\ \frac{\partial G}{\partial b_1} &= \sum_{i=1}^n 2(y_i - b_0 - b_1x_i) \cdot (-x_i) = 0\end{aligned}$$

Po úprave dostaneme systém normálnych rovníc:

$$\begin{aligned}nb_0 + b_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i, \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i\end{aligned}$$

Ak zapíšeme maticu plánu $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$, vektor odhadov parametrov B_0, B_1 $B = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$,

vektor meraní $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, potom môžeme písať systém normálnych rovníc v tvare maticovej rovnice

$$\begin{aligned} XB &= Y & / \cdot X^T \text{ zľava} & & X^T \text{ je transponovaná matica} \\ X^T XB &= X^T Y & & & X^T X \text{ je štvorcová matica, ak } \exists k \text{ nej} (X^T X)^{-1} \\ (X^T X)^{-1}(X^T X)B &= (X^T X)^{-1}X^T Y \\ B &= (X^T X)^{-1}X^T Y \end{aligned}$$

Parameter b_0 nazývame lokujúca konštanta a znamená hodnotu premennej Y za predpokladu, že premenná X sa rovná nule.

Parameter b_1 nazývame regresný koeficient, znamená prírastok (úbytok) premennej Y zodpovedajúci zmene premennej X o jednu jednotku. Regresný koeficient b_1 má vždy rovnaké znamienko ako koeficient korelácie r_{xy} . Regresná priamka prechádza bodom $[\bar{x}, \bar{y}]$.

Pri riešení regresnej úlohy prichádza do úvahy viacero regresných funkcií, preto sa skúma, ktorý z týchto modelov najlepšie vystihuje závislosť medzi premennými X a Y . To sa dá merať rôznymi charakteristikami: reziduálnym súčtom štvorcov, reziduálnym rozptylom, štandardnou odchýlkou rezíduí, koeficientom determinácie alebo preveriť rôznymi testami. Všetky tieto údaje nájdeme v tabuľke analýzy rozptylu, ktorá je výstupom analytického nástroja Regresia (Regression).

- Reziduálny súčet štvorcov $SSE = \sum_i (y_i - \hat{y}_i)^2$ – cieľom metódy najmenších štvorcov bolo minimalizovať jeho hodnotu, t.j. minimalizovať variabilitu nevysvetlenú regresným modelom.
- Reziduálny rozptyl je $s_{rez}^2 = \frac{SSE}{n-2}$.
- Štandardná odchýlka rezíduí $s_{rez} = \sqrt{s_{rez}^2}$.

Z dvoch regresných modelov, ktoré by teoreticky prichádzali do úvahy, je lepší na základe analýzy rezíduí ten, ktorý má menšie hodnoty SSE , s_{rez}^2 a s_{rez} .

Definícia 5.8 Koeficient (index) determinácie je definovaný podielom variability vysvetlenej modelom k celkovej variabilite $R^2 = \frac{SSR}{SSY} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$; $R^2 \in \langle 0, 1 \rangle$.

Koeficient determinácie ukazuje, koľko percent celkovej variácie (rozptylu) premennej Y je vysvetlenej modelom. Čím viac sa hodnota R^2 blíži k 1, tým je praktické použitie modelu rozumnejšie. Ak je hodnota R^2 pod 0,7, použitie regresného modelu má skôr len teoretický význam. Hodnoty nad 0,9 znamenajú „slušný“ model pre praktické využitie. Platí, že čím väčší je podiel variability SSR na celkovej variabilite SSY , tým je závislosť X a Y silnejšia, a preto je z dvoch modelov „kvalitnejší“ ten, ktorý má väčší koeficient determinácie R^2 . Koeficient

determinácie R^2 je rovný druhej mocnine korelačného koeficientu r_{xy} a vyjadruje stupeň príčinnej závislosti Y od X .

Príklad 5.6 Predpokladajme, že medzi náhodnými premennými X a Y existuje lineárna závislosť. Náhodne bolo vybraných 8 štatistických jednotiek, na ktorých sa skúmali hodnoty uvažovaných náhodných premenných X a Y . Namerané údaje sú zaznamenané v nasledujúcej tabuľke:

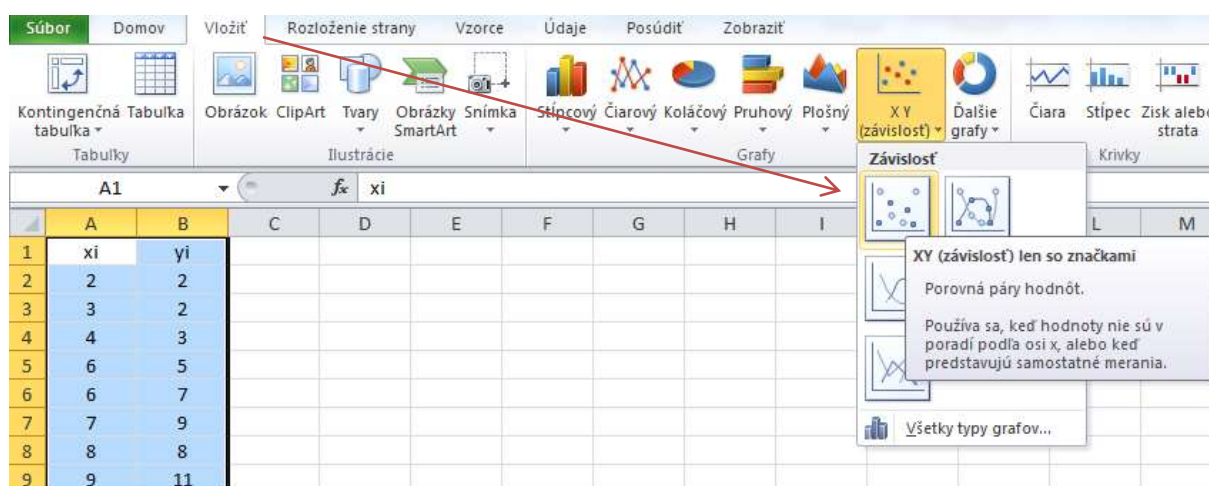
x_i	2	3	4	6	6	7	8	9
y_i	2	2	3	5	7	9	8	11

Zostrojte bodový diagram závislosti Y od X pre daný výberový súbor.

Cez zobrazené body preložte regresnú priamku, vložte do grafu rovnicu regresnej priamky, hodnotu koeficientu determinácie.

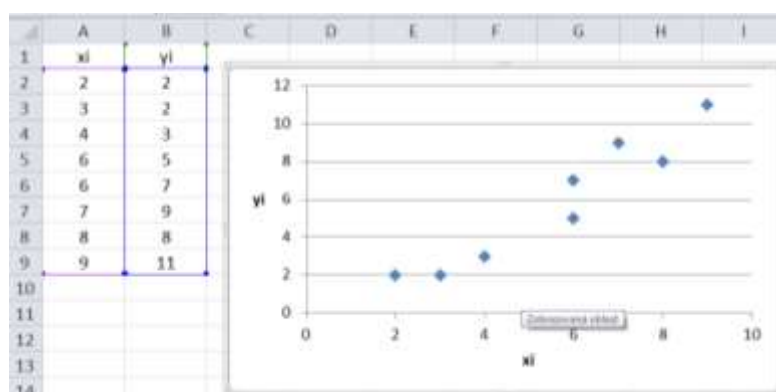
Vytvorte kovariančnú a korelačnú maticu (v Exceli pomocou analytických nástrojov koeficient korelácie a kovarianciu získame z korelačnej a kovariančnej matice).

Riešenie. Vizualne môžeme najlepšie posúdiť charakter a mieru vzájomnej závislosti náhodnej premennej Y od X z grafického vyjadrenia – z bodového grafu. Bodový graf umožňuje prezentovať rozdelenie hodnôt podľa dvoch premenných súčasne. Hodnoty prvej nezávislej premennej zobrazujeme na x -ovej osi a hodnoty druhej závislej premennej na y -ovej osi. V grafe je príslušná štatistická jednotka (príslušné pozorovanie) zobrazená bodom so súradnicami, zodpovedajúcimi hodnotám prvej premennej (X) a druhej premennej (Y). Pri vytváraní bodového grafu v Exceli volíme po označení vstupných údajov postupnosť príkazov Vložiť→Grafy→XY závislosť len so značkami (Insert→Chart→Scatterplot), obr. 5.21.



Obr. 5.21 Vytváranie bodového grafu nameraných hodnôt v Exceli

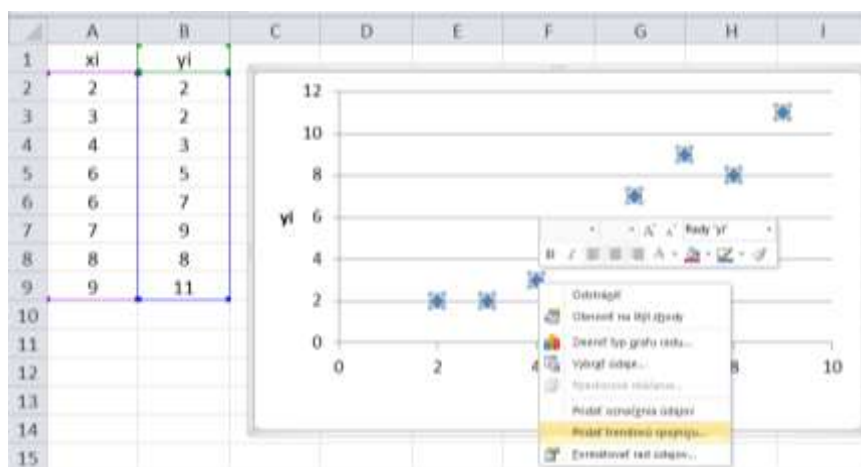
Výstupom je korelogram na obr. 5.22.



Obr. 5.22 Grafické znázornenie závislosti premennej Y od X – tzv. korelačný diagram

Z grafu vidíme, že s rastúcimi hodnotami x_i rastú aj hodnoty y_i – hovoríme o priamej závislosti, pričom z usporiadania bodov grafu možno predpokladať, že ide o závislosť lineárnu (body „ležia približne“ na priamke).

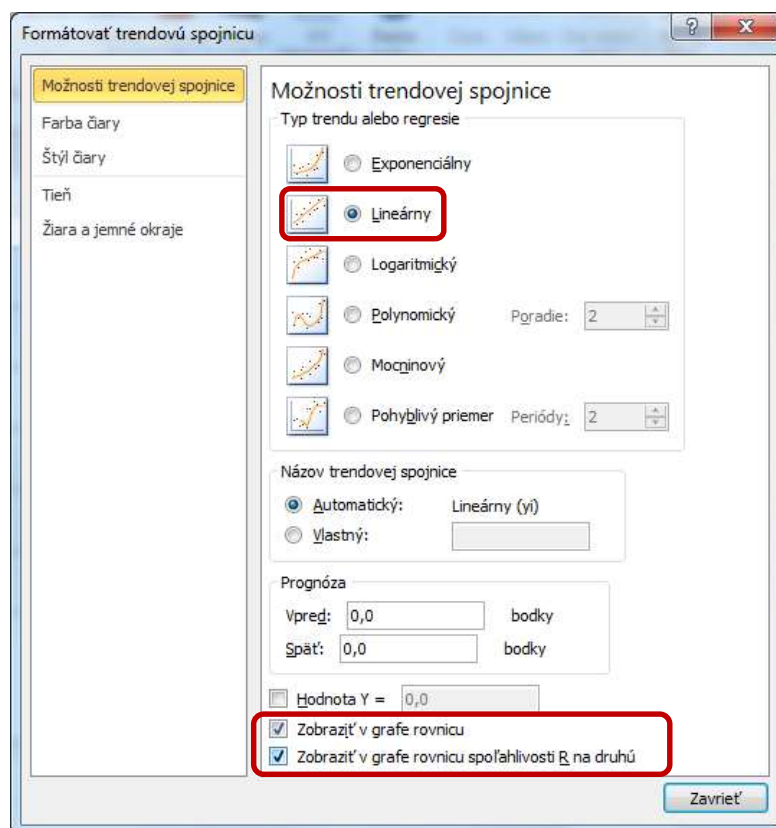
Ak chceme cez zobrazené body preložiť regresnú priamku, ako prvé označíme myšou body grafu, obr. 5.23. Následne pravým tlačidlom myši aktivujeme okno, v ktorom zvolíme Pridať trendovú spojnicu...



Obr. 5.23 Po označení sa všetky body grafu stanú aktívnymi

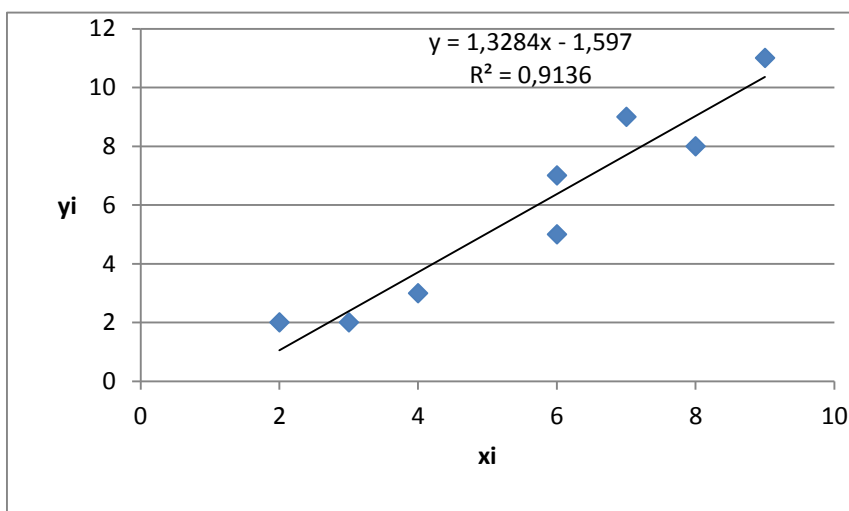
Dialógové okno nám pri zadávaní typu regresného modelu - trendovej spojnice (fitovacej krivky) ponúka možnosť modelovať závislosť skúmaných premenných pomocou priamky (lineárny model), exponenciálnej, logaritmickej, mocnínovej funkcie či polynómu.

Tento výber kriviek je tiež výhodou grafického riešenia regresie, nakoľko analytický nástroj Regression v Exceli umožňuje spracovávať iba také závislosti, ktoré sú v parametroch lineárne, resp. linearizovateľné (transformáciou, napr. logaritmovaním, substitúciou...). Pomocou ponuky Zobraziť v grafe rovnicu a Zobraziť v grafe rovnicu spoľahlivosti R na druhú môžeme do bodového grafu vpísať tiež rovnicu regresnej krivky a index determinácie R^2 , obr. 5.24.



Obr. 5.24 Dialógové okno, ktoré sa zobrazí pri formátovaní regresnej funkcie

Po výbere a potvrdení našich požiadaviek dostaneme výstup na obr. 5.25.

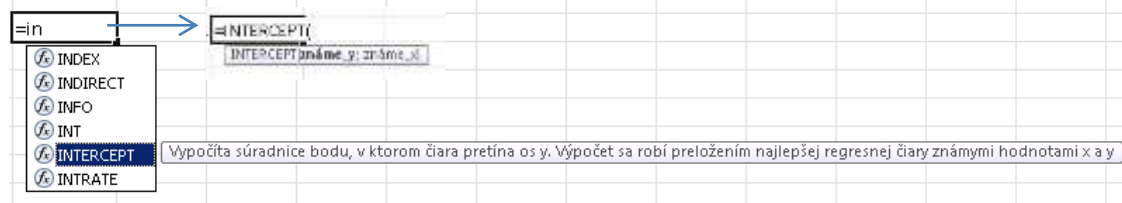


Obr. 5.25 Bodový graf s regresnou priamkou obsahuje aj jej predpis a index determinácie

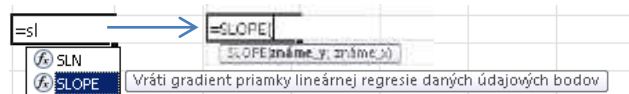
Hľadaná rovnica regresnej priamky $y = -1,597 + 1,328x$ umožňuje odhadovať hodnoty premennej Y , ak poznáme hodnotu premennej X (umožňuje tzv. predikciu). Pritom hodnota koeficientu determinácie R^2 hovorí, nakoľko je nami nájdený model „dobrý“, t.j. vyjadruje, aký podiel variability premennej Y regresný model vysvetľuje. V našom prípade 91,36% variability premennej Y vieme vyjadriť lineárnym vzťahom s premennou X .

Poznámka 5.4 Na výpočet parametrov regresnej priamky môžeme použiť aj preddefinované štatistické funkcie.

Číselnú hodnotu lokujúcej konštanty b_0 vypočítame funkciou `=INTERCEPT(známe_y; známe_x)`, pričom je dôležité dodržiavať poradie premenných v požadovanom tvare.

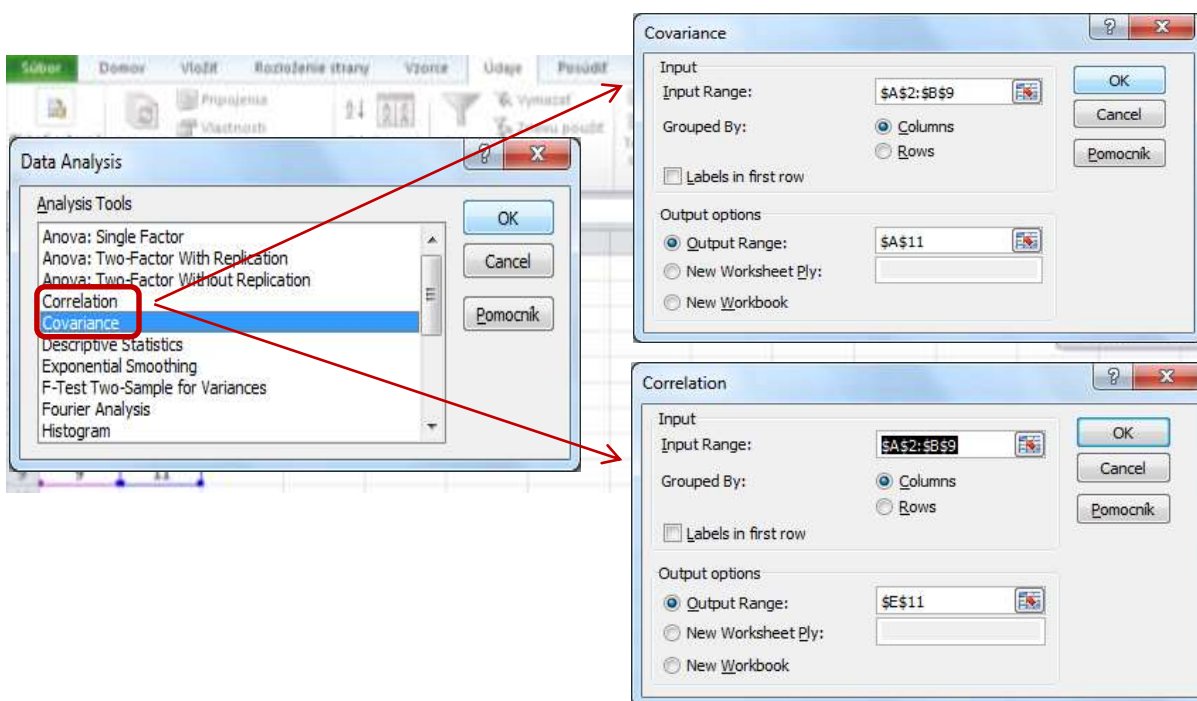


Regresný koeficient b_1 vypočíta funkcia `=SLOPE(známe_y; známe_x)`.



Rovnicu regresnej priamky dostaneme dosadením vypočítaných koeficientov do predpisu $y = b_0 + b_1x$.

Kovariačnú a korelačnú maticu vieme v Exceli vytvoriť pomocou analytických nástrojov Kovariancia (Covariance) a Korelácia (Correlation). Aktivujeme ich postupnosťou príkazov Údaje→Analýza dát→Kovariancia resp. Korelácia (Data→Data analysis→ Covariance resp. Correlation). Ich dialógové okná vyžadujú špecifikáciu Vstupnej oblasti: (Input Range), t.j. označenie oblasti, v ktorej sa nachádzajú empirické hodnoty premenných X a Y . Ďalej je potrebné uviesť, či sú zapísané v riadkoch (Rows) alebo v stĺpcoch (Columns), a potvrdiť popisy premenných, nachádzajúce sa v prvom riadku (Labels in First Row). V časti Možnosti výstupu (Output options) špecifikujeme umiestnenie výstupu. Ich náhľad je na obr. 5.26.



Obr. 5.26 Aktivne okná analytických nástrojov Correlation a Covariance

Výstupom je postupné zobrazenie kovariančnej a korelačnej matice, pričom obe matice sú uvedené v tvare dolného trojuholníka (hodnoty v hornom trojuholníku sú zhodné s hodnotami v dolnom trojuholníku), obr. 5.27.

	A	B	C	D	E	F	G
1	x_i	y_i					
2	2	2					
3	3	2					
4	4	3					
5	6	5					
6	6	7					
7	7	9					
8	8	8					
9	9	11					
10							
11		Column 1	Column 2			Column 1	Column 2
12	Column 1	5,23438			Column 1	1	
13	Column 2	6,95313	10,1094		Column 2	0,95584	1

Obr. 5.27 Zobrazenie kovariančnej a korelačnej matice

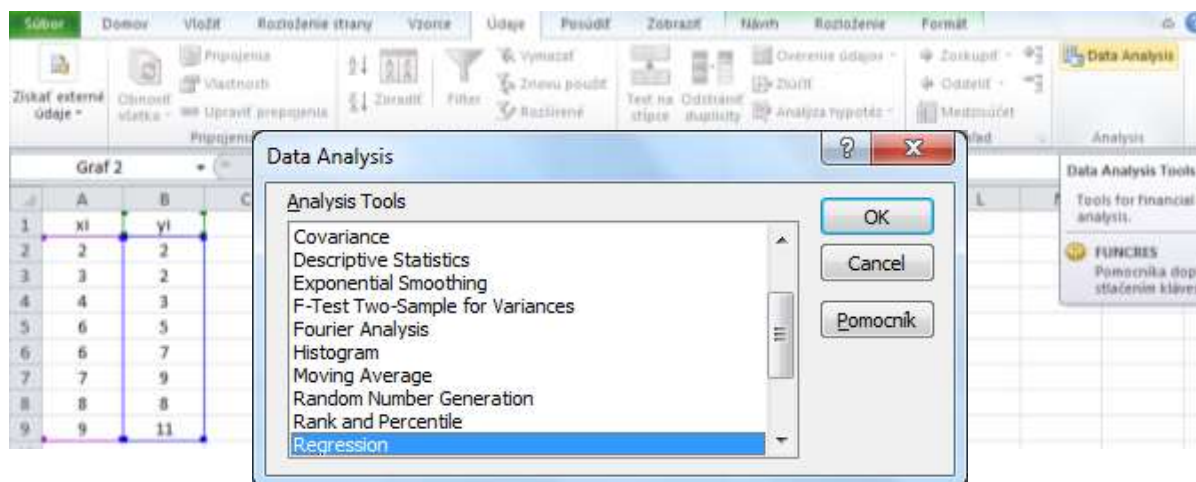
Poznámka 5.5 Kovarianciu medzi hodnotami dvoch premenných počítame zriedkavo. Častejšie sa používa kovariančná matica, ak chceme charakterizovať vzájomnú lineárnu závislosť medzi všetkými dvojicami ukazovateľov pri analýze hodnôt viac premenných. Kovariančná matica je symetrická podľa diagonály, nakoľko kovariancia medzi prvým a druhým ukazovateľom (premennou) je taká istá ako medzi druhým a prvým ukazovateľom atď. Na diagonále kovariančnej matice sú rozptyly náhodnej premennej, ktorá sa nachádza v danom riadku matice. Analogicky rovnaké vlastnosti má aj korelačná matica, ktorá sa skladá z vypočítaných hodnôt koeficientov korelácie medzi všetkými možnými dvojicami premenných. V prvom riadku sú hodnoty koeficientov korelácie medzi hodnotami prvej premennej s hodnotami ostatných premenných v druhom riadku medzi hodnotami druhej premennej a hodnotami ostatných premenných atď. Korelačná matica je tiež symetrická. Koeficient korelácie medzi hodnotami prvej premennej a hodnotami druhej premennej je taký istý, ako je koeficient korelácie medzi hodnotami druhej premennej a hodnotami prvej premennej. Na diagonále korelačnej matice sú jednotky, nakoľko koeficient korelácie medzi hodnotami nejakej premennej a hodnotami tej istej premennej sa rovná vždy 1 – všetky body teda ležia na priamke. Čím je korelačná matica väčšieho rozsahu, tým je jej použitie efektívnejšie. Analýzou kovariančnej matice vieme porovnať, medzi ktorými dvojicami je najväčšia závislosť, všimame si pritom, medzi ktorými je závislosť kladná (priama), medzi ktorými je záporná (nepriama) a tiež medzi ktorými dvojicami závislosť nie je.

V bunke B13 je kovariancia $cov(x, y) = 6,95313$, čo interpretujeme ako priamu lineárnu závislosť medzi premennými X a Y .

V bunke F13 je koeficient korelácie $r_{xy} = 0,95584$, čo interpretujeme ako silnú priamu lineárnu závislosť medzi premennými X a Y .

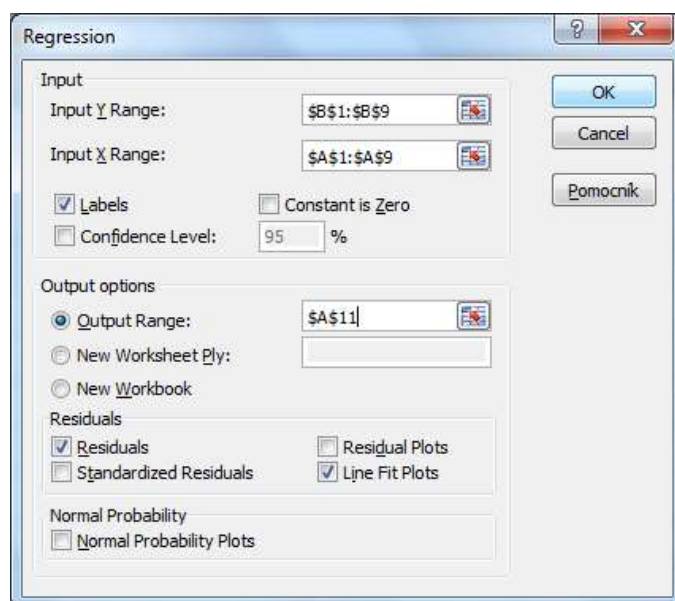
Príklad 5.7 Riešte regresnú úlohu z predchádzajúceho zadania príkladu pomocou nástroja Regresia v Analýze dát.

Riešenie. Analytický nástroj, ktorým môžeme riešiť v Exceli regresnú úlohu, aktivujeme postupnosťou Údaje→Analýza dát→Regresia (Data→Data analysis→Regression), obr. 5.28.



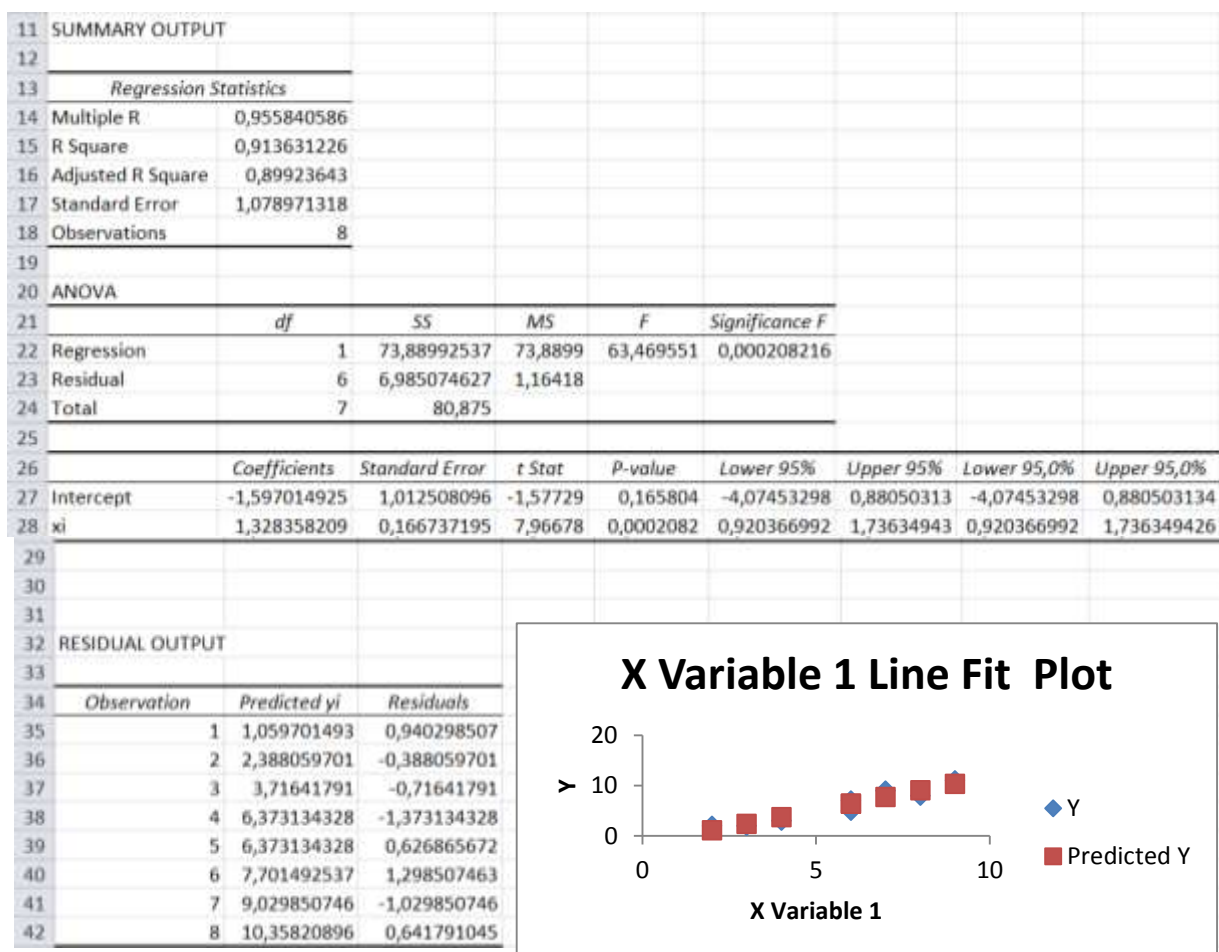
Obr. 5.28 Výber analytického nástroja Regression

Dialógové okno nástroja Regression vyžaduje špecifikovať Vstupnú oblasť (Input) v tvare osobitného označenie oblasti, v ktorej sa nachádzajú namerané hodnoty závislej premennej Y (Input Y Range) a osobitného označenie oblasti, v ktorej sa nachádzajú namerané hodnoty nezávislej premennej X (Input X Range). Popisy premenných v prvej bunke označenej oblasti potvrdzujeme v Labels. V časti Možnosti výstupu (Output options) špecifikujeme umiestnenie výstupu (Output Range), volíme výpis rezíduí (Residuals) a zobrazenie tzv. vyrovnaných hodnôt. Vzhľadom na zadanie úlohy sú naše voľby postačujúce. Náhľad dialógového okna je na nasledujúcom obr. 5.29.



Obr. 5.29 Dialógové okno analytického nástroja Regression

Výstupom je tabuľka, skladajúca sa z niekoľkých častí (tabuliek) a bodový graf nameraných hodnôt (Y) a modelovaných hodnôt (Predicted Y), prislúchajúcich hodnotám nezávislej premennej X, obr. 5.30.

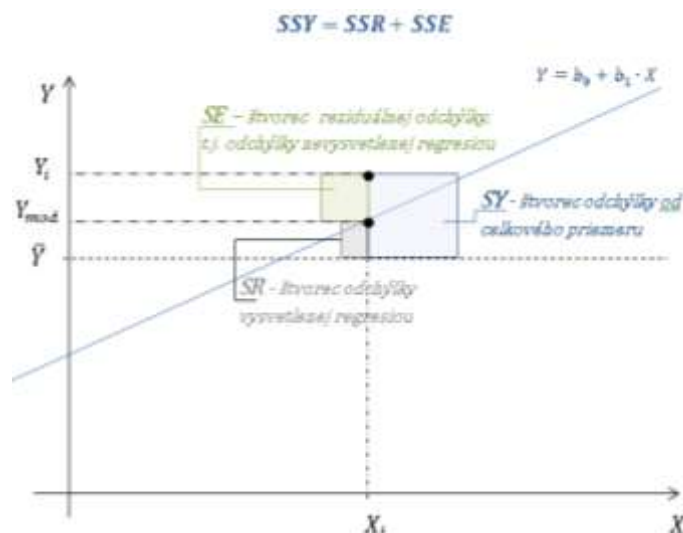


Obr. 5.30 Výstup Regresie v Analýze dát

Prvá tabuľka Regresná štatistika (Regression Statistics) obsahuje absolútnu hodnotu koeficientu korelácie $|r|$ (Multiple R), index determinácie R^2 (R Square), upravený (korigovaný) koeficient determinácie $r_{adj}^2 = 1 - (1 - r^2) \frac{n-1}{n-2}$ (Adjusted R Square), štandardnú

odchýlku rezíduí $s_{rez} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{SSE}{n-2}}$ (Standard Error), rozsah VS n (Observations).

Ďalšou časťou výstupu je tabuľka analýzy rozptylu ANOVA– ANalysis Of VAriance, ktorá dáva odpoveď na otázku, či je regresný model štatisticky vhodný. V prvom stĺpci tabuľky df sú uvedené počty stupňov voľnosti. V stĺpci SS sú uvedené súčty štvorcov odchýlok zo vzťahu $\sum(y_i - \hat{y}_i)^2 + \sum(\hat{y}_i - \bar{y})^2 = \sum(y_i - \bar{y})^2$, ktorý môžeme zapísať skrátené v tvare $SSR + SSE = SSY$, kde SSR – SUM OF SQUARES DUE TO REGRESSION – variabilita, vysvetlená regresným modelom, SSE – SUM OF SQUARES DUE TO ERROR – variabilita, nevysvetlená regresným modelom (reziduálny súčet štvorcov), SSY – SUM OF SQUARES TOTAL je celková variabilita premennej Y. Na obr. 5.31 je tento rozklad zobrazený pre prípad i -tého pozorovania.



Obr. 5.31 Rozklad štvorca celkovej variability na zložky reziduálnu a regresnú

Stĺpec MS obsahuje hodnotu priemerného štvorca regresie MSR (Mean Square of Regression) a priemerného štvorca chýb MSE (Mean Square of Errors).

Pre interpretáciu ďalších údajov z tabuľky analytických nástrojov sa teoretické základy nachádzajú v kapitole Vyhodnocovanie experimentov.

Celkový F -test, ktorým testujeme nulovú hypotézu H_0 : Lineárny model nie je štatisticky významný. Testovacia štatistika má Fisherovo rozdelenie s $(1, n - 2)$ stupňami voľnosti. Ak je hodnota Významnosť F (p -hodnota) (Significance F) v poslednom stĺpci menšia ako zvolená hladina významnosti α , model je štatisticky vhodný.

	stupne voľnosti	SS	MS	F	p - hodnota
regresia	1	SSR	$MSR = SSR/1$	hodnota testovacej štatistiky $F = \frac{MSR}{MSE}$	pravdepodobnosť chyby, ktorej sa dopustíme, keď zamietneme H_0 , ktorá platí
reziduá	$n - 2$	SSE	$s_{rez}^2 = MSE = SSE/(n - 2)$		
celkom	$n - 1$	SSY			

Tretia časť tabuľky obsahuje v stĺpci Coefficients parametre regresnej priamky (zhodujú sa s hodnotami vypočítanými štatistickými funkciami INTERCEPT a SLOPE) a t -test pre nulovosť koeficientov, ktorým testujeme $H_0: b_i = 0$. Testovacia štatistika t má Studentovo rozdelenie s $n - 2$ stupňami voľnosti.

	koeficienty	s	t	p - hodnota	intervaly spoľahlivosti (IS)
lokujúca konštanta	b_0	$s(b_0)$	hodnota testovacej štatistiky $t = \frac{b_0}{s(b_0)}$	$p(b_0)$	IS pre b_0
regresný koeficient	b_1	$s(b_1)$	hodnota testovacej štatistiky $t = \frac{b_1}{s(b_1)}$	$p(b_1)$	IS pre b_1

Poslednou časťou je nami vybraná tabuľka predikovaných hodnôt (pre každý prvok x_i výberového súboru je vypočítaná v rámci modelu tzv. vyrovnaná – očakávaná – modelovaná hodnota $\hat{y}_i = y_{i\ mod}$) a rezíduí $e_i = y_i - \hat{y}_i$.

Poznámka 5.6 Platí $SSE = \sum_i e_i^2$.

Z uvedených tabuliek pre riešenie nášho príkladu vyplýva:

Z prvej tabuľky v riadku 14 prečítame absolútnu hodnotu koeficientu korelácie $|r_{xy}| = 0,955840586$. Z grafického zobrazenia (vidíme, že závislosť je priama, pretože s rastúcimi hodnotami x_i rastú i hodnoty y_i) alebo z porovnania s regresným koeficientom (regresný koeficient má rovnaké znamienko ako korelačný koeficient) môžeme písať $r_{xy} = 0,955840586$, čo interpretujeme ako priamu silnú (tesnú) lineárnu závislosť medzi premennými X a Y .

V riadku 15 sa nachádza index (koeficient) determinácie $R^2 = \frac{SSR}{SSY} = 0,913631226$, ktorý udáva podiel variability vysvetlenej regresným modelom, tzn. 91,36% variability premennej Y sa dá vysvetliť lineárnym vzťahom s premennou X . Táto hodnota ilustruje praktickú cennosť modelu.

Štandardná odchýlka rezíduí s_{rez} v riadku 17 má hodnotu 1,078971318, čo znamená, že skutočné hodnoty premennej Y sa odchyľujú od hodnôt, ležiacich na regresnej priamke (predikovaných hodnôt) približne o $\pm 1,078971318$.

Keďže v druhej časti, v tabuľke ANOVA v riadku 22 je p -hodnota (Významnosť F , Significance F) rovná 0,000208216, testovanú hypotézu H_0 o nezávislosti pozorovaných premenných X a Y neprijímame (zamietame). Pravdepodobnosť chyby, ktorej sa týmto dopustíme, je takmer nulová. To znamená, že navrhnutý lineárny model je štatisticky významný.

Koeficienty regresnej priamky nájdeme v riadkoch 27 a 28 nasledujúcej tabuľky. Lokujúca konštanta $b_0 = -1,597014925$ a smernica priamky, t.j. regresný koeficient $b_1 = 1,328358209$. Odtiaľ bodový odhad regresnej priamky: $y = -1,597 + 1,328x$.

Testovať lineárnu nezávislosť X a Y môžeme aj pomocou t -testu, ktorý je založený na tom, že regresný koeficient b_1 ako smernica priamky vyjadruje zmenu Y pri zmene X o jednu jednotku. Ak $b_1 = 0$, regresná priamka je rovnobežná s x -ovou osou, a teda aj po zmene nezávislej premennej X sa nemenia hodnoty Y , a preto nemôžeme hovoriť o lineárnej závislosti X a Y . Testujeme $H_0: b_1 = 0 >< H_1: b_1 \neq 0$. V riadku 28 tabuľky sú uvedené hodnota testovacej štatistiky, p -hodnota a 95%-ný interval spoľahlivosti pre koeficient b_1 . Porovnaním hodnoty testovacej štatistiky $t(b_1) = 7,966778$ s kvantilom $t_6^{kv}(0,975) = T.INV(0,975; 6) = 2,446912$ alebo na základe p -hodnoty $p(b_1) = 0,0002082 < 0,05 = \alpha$ alebo tiež na základe toho, že nula nepatrí do intervalu $(0,920367; 1,73634943)$, môžeme prijať záver, že pri zvolenej hladine významnosti $\alpha = 0,05$ zamietame nulovú hypotézu o tom, že premenné X a Y sú lineárne nezávislé a prijímame alternatívnu hypotézu o lineárnom vzťahu medzi X a Y .

Príklad 5.8 Odhadnite regresnú priamku z príkladu 5.6 pomocou tzv. systému normálnych rovníc.

Riešenie. Neznáme parametre b_0, b_1 regresnej priamky v sústave normálnych rovníc

$nb_0 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$
 $b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$ je možné podľa Cramerovho pravidla za podmienky $D = \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} \neq 0$, nájsť v tvare $b_0 = \frac{D_1}{D}$, $b_1 = \frac{D_2}{D}$, kde $D_1 = \begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix}$ a $D_2 = \begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix}$.
 Riešenie v Exceli je na obr. 5.32.

	A	B	C	D	E	F	G	H	I
1	xi	yi	sustava normalnych rovníc: $b_0 \cdot n + b_1 \cdot \sum x_i = \sum y_i$ $b_0 \cdot \sum x_i + b_1 \cdot \sum x_i^2 = \sum x_i y_i$						
2	2	2							
3	3	2							
4	4	3							
5	6	5	n= =COUNT(B2:B9)						
6	6	7	Σxi= =SUM(A2:A9)						
7	7	9	Σxi^2= =SUMSQ(A2:A9)						
8	8	8	Σyi= =SUM(B2:B9)						
9	9	11	Σxiyi= =SUMPRODUCT(A2:A9;B2:B9)						
10									
11	D=	=D5	=D6	D1=	=D8	=D6	D2=	=D5	=D8
12		=D6	=D7		=D9	=D7		=D6	=D9
13	D=	=MDETERM(B11:C12)		D1=	=MDETERM(E11:F11)		D2=	=MDETERM(H11:I12)	
14									
15	b0=	=E13/\$B\$13							
16	b1=	=H13/\$B\$13							
17	odhad regresnej priamky:								
18	y=	=B15	+	=B16	*x				

	A	B	C	D	E	F	G	H	I
1	xi	yi							
2	2	2			systava normalnych rovníc: $b_0 \cdot n + b_1 \cdot \sum x_i = \sum y_i$ $b_0 \cdot \sum x_i + b_1 \cdot \sum x_i^2 = \sum x_i y_i$				
3	3	2							
4	4	3							
5	6	5	n= 8						
6	6	7	Σxi= 45						
7	7	9	Σxi^2= 295						
8	8	8	Σyi= 47						
9	9	11	Σxiyi= 320						
10									
11	D=	8	45	D1=	47	45	D2=	8	47
12		45	295		320	295		45	320
13	D=	335		D1=	-535		D2=	445	
14									
15	b0=	-1,597							
16	b1=	1,32836							
17	odhad regresnej priamky:								
18	y=	-1,59701	+	1,328358	*x				

Obr. 5.32 Výpočet koeficientov regresnej priamky z normálneho systému rovníc použitím Cramerovho pravidla – riešenie v Exceli

Ak zapíšeme systém normálnych rovníc v maticovom tvare $AB = C$, kde $A = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$; $B = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$; $C = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$, potom vektor neznámych koeficientov $B = A^{-1} C$.

Riešenie systému normálnych rovníc pomocou maticových rovníc v Exceli je naznačené na nasledujúcom obr. 5.33.

	A	B	C	D	E	F	G	H	I
1	xi	yi			sustava normalnych rovníc:				
2	2	2			$b_0 \cdot n + b_1 \cdot \sum x_i = \sum y_i$				
3	3	2			$b_0 \cdot \sum x_i + b_1 \cdot \sum x_i^2 = \sum x_i y_i$				
4	4	3							
5	6	5		n= 8					
6	6	7		$\sum x_i = 45$					
7	7	9		$\sum x_i^2 = 295$					
8	8	8		$\sum y_i = 47$					
9	9	11		$\sum x_i y_i = 320$					
10									
11	A=	=D5	=D6		B=	b0		C=	=D8
12		=D6	=D7			b1			=D9
13									
14	invA=	=MINVERSE(B11:C12)			B = invA*C =	=MMULT(B14:C15;I11:I12)			
15									
16									
17	odhad regresnej priamky:								
18	y=	=F14	+	=F15	*x				

	A	B	C	D	E	F	G	H	I
1	xi	yi			sustava normalnych rovníc:				
2	2	2			$b_0 \cdot n + b_1 \cdot \sum x_i = \sum y_i$				
3	3	2			$b_0 \cdot \sum x_i + b_1 \cdot \sum x_i^2 = \sum x_i y_i$				
4	4	3							
5	6	5		n= 8					
6	6	7		$\sum x_i = 45$					
7	7	9		$\sum x_i^2 = 295$					
8	8	8		$\sum y_i = 47$					
9	9	11		$\sum x_i y_i = 320$					
10									
11	A=	8	45		B=	b0		C=	47
12		45	295			b1			320
13									
14	invA=	0,880597	-0,13433		B = invA*C =	-1,597			
15		-0,134328	0,023881			1,3284			
16									
17	odhad regresnej priamky:								
18	y=	-1,59701	+	1,328358	*x				

Obr. 5.33 Výpočet koeficientov regresnej priamky z normálneho systému rovníc riešením maticovej rovnice v Exceli

Ak lineárny regresný model zapíšeme pomocou matice plánu experimentu $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$,

potom z riešenia normálneho systému v maticovom tvare pre odhad vektora parametrov $B = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$ platí $B = (X^T X)^{-1} X^T Y$, kde X^T je transponovaná matica. Riešenie v Exceli môžeme vidieť na nasledujúcich obr. 5.34.

	A	B	C	D	E	F	G	H	I
10									
11	X=	1	2		B=	b0		Y=	2
12		1	3			b1			2
13		1	4						3
14		1	6						5
15		1	6						7
16		1	7						9
17		1	8						8
18		1	9						11
19									
20	XT=	=TRANSPOSE(B11:C18)							
21									
22									
23	XT*X=	=MMULT(B20:I21;B11:C18)			inv(XT*X)=	=MINVERSE(B23:C24)			
24									
25									
26	inv(XT*X)*XT=	=MMULT(F23:G24;B20:I21)							
27									
28									
29	B=inv(XT*X)*XT*Y=	=MMULT(B26:I27;I11:I18)							
30									

19									
20	XT=	1	1	1	1	1	1	1	1
21		2	3	4	6	6	7	8	9
22									
23	XT*X=	8	45		inv(XT*X)=	0,8806	-0,13433		
24		45	295			-0,1343	0,023881		
25									
26	inv(XT*X)*XT=	0,6119403	0,477612	0,343284	0,07463	0,07463	-0,0597	-0,19403	-0,3284
27		-0,0865672	-0,062687	-0,038806	0,00896	0,00896	0,032836	0,056716	0,0806
28									
29	B=inv(XT*X)*XT*Y=	-1,5970149							
30		1,32835821							

Obr. 5.34 Výpočet koeficientov regresnej priamky vyjadrenej pomocou matice plánu experimentu ak normálny systém rovníc riešime v maticovom tvare - riešené v Exceli

Skrátene môžeme tiež zapísať:

	A	B	C	D	E	F	G	H	I	J	K	L
10												
11	X=	1	2		B=	b0		Y=	2			
12		1	3			b1			2			
13		1	4						3			
14		1	6						5			
15		1	6						7			
16		1	7						9			
17		1	8						8			
18		1	9						11			
19												
20	B=inv(XT*X)*XT*Y=	=MMULT(MMULT(MINVERSE(MMULT(TRANSPOSE(B11:C18);B11:C18));TRANSPOSE(B11:C18));I11:I18)										
21												

	A	B	C	D	E	F	G	H	I
10									
11	X=	1	2		B=	b0		Y=	2
12		1	3			b1			2
13		1	4						3
14		1	6						5
15		1	6						7
16		1	7						9
17		1	8						8
18		1	9						11
19									
20	B=inv(XT*X)*XT*Y=		-1,59701						
21			1,328358						

Obr. 5.35 Výpočet koeficientov regresnej priamky vyjadrenej pomocou matice plánu experimentu, ak normálny systém rovníc riešime v maticovom tvare - skrátené riešenie v Exceli

Príklad 5.9 Do vodnej nádrže unikla jedovatá látka. Pre likvidáciu jej účinku bol aplikovaný neutralizačný prostriedok. Od okamihu jeho aplikácie sa niekoľkokrát merala koncentrácia jedu. Výsledky merania sú uvedené v nasledujúcej tabuľke:

doba (min)	5	12	20	26	29	40	65	135
koncentrácia (‰)	20	18	18	17	17	15	14	5

Za predpokladu približne lineárneho poklesu koncentrácie jedovatej látky určte:

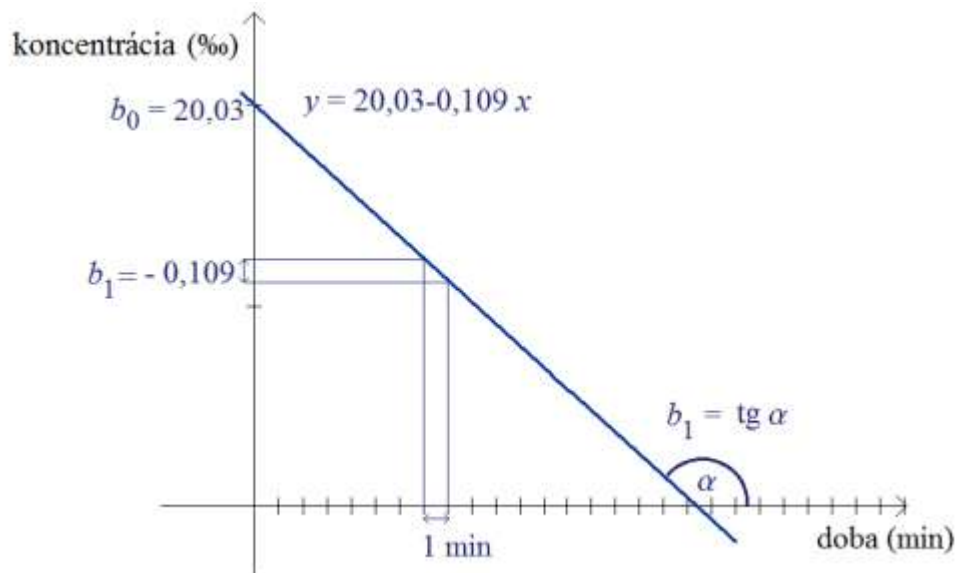
- koeficienty regresnej priamky a na jej grafe zobrazte ich význam (predpoklad lineárnej závislosti overte pomocou p -hodnoty testu linearity a testu pre korelačný koeficient);
- koncentráciu jedu po 1 hodine a po 100 minútach;
- okamih, kedy bude koncentrácia jedovatej látky nulová.

Riešenie. Najefektívnejší spôsob výpočtu v Exceli je výpočet pomocou analytického nástroja Regression. Jeho výstup je na obr. 5.36.

	A	B	C	D	E	F	G
10							
11	SUMMARY OUTPUT						
12							
13	<i>Regression Statistics</i>						
14	Multiple R	0,991549					
15	R Square	0,983169					
16	Adjusted R Square	0,980364					
17	Standard Error	0,648676					
18	Observations	8					
19							
20	ANOVA						
21		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
22	Regression	1	147,4753196	147,475	350,481	1,5E-06	
23	Residual	6	2,524680369	0,42078			
24	Total	7	150				
25							
26		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
27	Intercept	20,0335	0,333524735	60,066	1,4E-09	19,217395	20,84960619
28	doba(min.)	-0,109241	0,00583517	-18,721	1,5E-06	-0,1235191	-0,09496283

Obr. 5.36 Výstup analytického nástroja Regression

Odhady parametrov regresnej priamky sa nachádzajú v bunkách B27 a B28. Lokujúca konštanta $b_0 = 20,0335$ je uvedená v bunke B27 a smernica regresnej priamky, t.j. regresný koeficient $b_1 = -0,109241$ v bunke B28. Odtiaľ bodový odhad regresnej priamky je $y = 20,0335 - 0,109241x$. Jej graf je na nasledujúcom obr. 5.37.



Obr. 5.37 Graf regresnej priamky $y = 20,0335 - 0,109241x$

Lokujúca konštanta $b_0 = 20,0335$ určuje priesečník regresnej priamky s y-ovou osou. Interpretácia: V rámci nášho modelu bola koncentrácia jedu vo vodnej nádrži v čase úniku (t.j. v čase 0 min) 20,0335 ‰ (nie vždy má táto konštanta logický zmysel).

Regresný koeficient $b_1 = 2,26$ je smernicou regresnej priamky (je tangensom uhla, ktorý priamka zviera s kladnou x -ovou poloosou).

Interpretujeme: Ak sa doba pôsobenia neutralizačnej látky zvýši o 1 minútu, potom koncentrácia jedu v nádrži klesne o 0,109 ‰.

Absolútna hodnota koeficientu korelácie $|r_{xy}| = 0,991549$ je v bunke B14. Z grafického zobrazenia vidíme, že závislosť je nepriama, pretože s rastúcim hodnotami x_i klesajú hodnoty y_i . Tiež z porovnania s regresným koeficientom (regresný koeficient má rovnaké znamienko ako korelačný koeficient) môžeme písať $r_{xy} = -0,991549$, čo interpretujeme ako nepriamu silnú (tesnú) lineárnu závislosť medzi premennými X a Y .

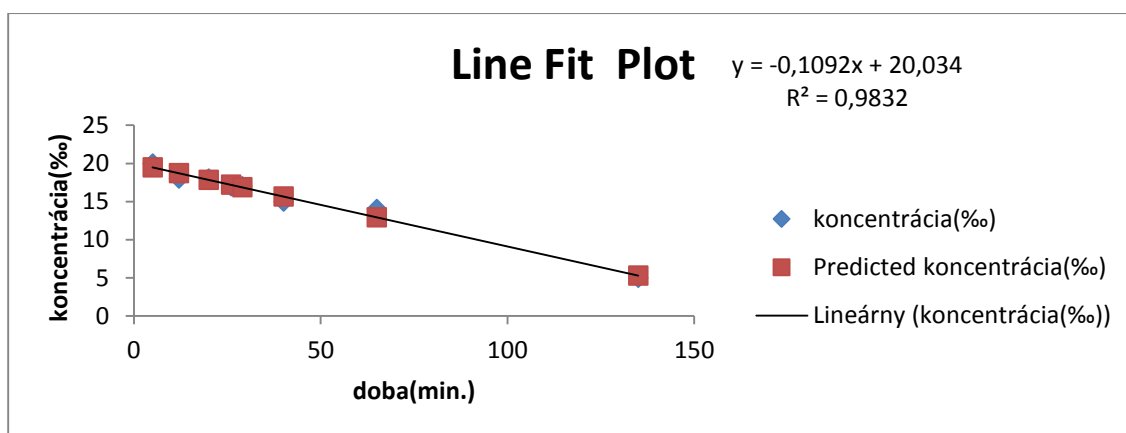
Na základe hodnoty koeficientu determinácie $R^2 = \frac{SSR}{SSY} = 0,9983169$, uvedenej v bunke B15, môžeme konštatovať, že 99,83% variability koncentrácie jedu v nádrži sa dá vysvetliť lineárnym vzťahom s dobou pôsobenia neutralizačnej látky. Iba 1,17% variability koncentrácie jedu zostalo regresným modelom nevysvetlenej, preto lineárny regresný model bol zvolený vhodne.

Neskreslený odhad koeficientu determinácie v základnom súbore je číslo 0,980364, nachádzajúce sa v bunke B16.

Pre celkový test linearity (F -test) je v bunke F22 uvedená p -hodnota $1,5 \cdot 10^{-6}$, čo je veľmi malé číslo, preto vo všetkých bežných hladinách významnosti zamietame nulovú hypotézu a prijímame alternatívnu hypotézu, že daný model je štatisticky významný, t.j. koncentrácia jedu a doba pôsobenia neutralizačnej látky sú lineárne závislé.

V bunke E28 je pre t -test štatistickej významnosti korelačného koeficientu b_1 uvedená rovnaká p -hodnota $1,5 \cdot 10^{-6}$, preto vo všetkých bežných hladinách významnosti prijímame alternatívnu hypotézu, že koncentrácia jedu a doba pôsobenia neutralizačnej látky sú lineárne závislé.

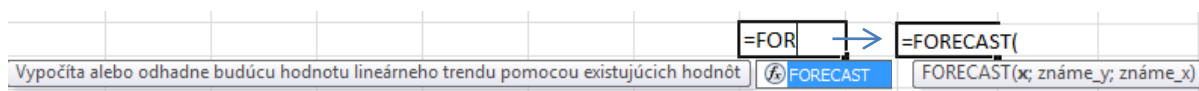
Súčasťou výstupu nástroja Regression je aj spoločný graf pôvodných nameraných hodnôt (koncentrácia) a model odhadovaných hodnôt (Predicted koncentrácia), obr. 5.38. Graf umožňuje intuitívne posúdiť kvalitu regresného modelu. Vidíme vysokú mieru zhody bodov reprezentujúcich skutočnosť a namodelovanej regresnej priamky.



Obr. 5.38 Na grafe Line Fit Plot vidíme skutočné (namerané) hodnoty a modelované hodnoty ležiace na regresnej priamke

Regresnú priamku $y = 20,0335 - 0,109241x$ považujeme za bodový odhad strednej hodnoty závislej premennej Y (koncentrácie jedu) a môžeme ju použiť na bodový odhad hodnoty y_i (koncentrácie) pre jednu konkrétnu hodnotu x_i (konkrétny čas). Na základe toho môžeme očakávať, že koncentráciu jedu vo vodnej nádrži po 1 hodine bude $y(60) = 20,0335 - 0,109241 \cdot 60 = 13,479042$ ‰ a koncentrácia jedu po 100 minútach $y(100) = 20,0335 - 0,109241 \cdot 100 = 9,1094028$ ‰.

Na odhad hodnôt závislej premennej Y (koncentrácie jedovatej látky) v Exceli môžeme pri modelovaní lineárnej závislosti použiť aj preddefinovanú funkciu $=FORECAST(x; známe_y; známe_x)$, ktorej argumenty uvádzame v poradí: hodnota času, pre ktorú chceme koncentráciu vypočítať; pole hodnôt závislej premennej – v našom prípade koncentrácie; pole hodnôt nezávislej premennej – v našom prípade času.



Dobu, v ktorej bude koncentrácia jedovatej látky nulová, určíme z regresnej priamky spätným vyjadrením premennej X : $x(0) = \frac{20,0335-0}{0,109241} = 183,38815$ min.

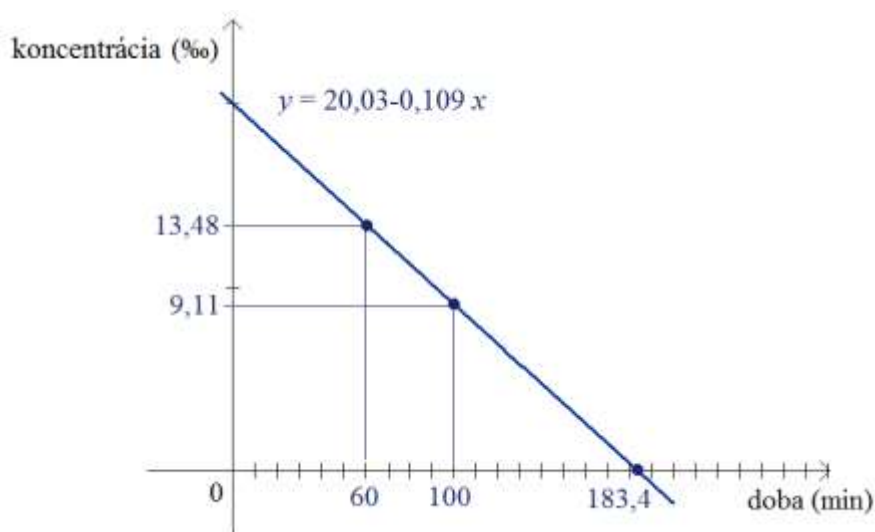
Výpočet v Exceli je na nasledujúcom obr. 5.39.

	A	B	C	D	E	F	G
1	doba(min.)	koncentrácia(‰)		Príklad. Likvidácia jedu vo vodnej nádrži.			
2	5	20		odhad regresnej priamky:			
3	12	18		$y = 20,0335 - 0,10924098 \cdot x$			
4	20	18					
5	26	17		t=60 min	y=	=FORECAST(60;B2:B9;A2:A9)	
6	29	17				=E3+F3*60	
7	40	15		t=100 min	y=	=FORECAST(100;B2:B9;A2:A9)	
8	65	14				=E3+F3*100	
9	135	5		y=0‰	x=	=(E3-0)/-F3	

	A	B	C	D	E	F	G
1	doba(min.)	koncentrácia(‰)		Príklad. Likvidácia jedu vo vodnej nádrži.			
2	5	20		odhad regresnej priamky:			
3	12	18		$y = 20,0335 - 0,10924098 \cdot x$			
4	20	18					
5	26	17		t=60 min	y=	13,479042	
6	29	17				13,479042	
7	40	15		t=100 min	y=	9,1094028	
8	65	14				9,1094028	
9	135	5		y=0‰	x=	183,38815	

Obr. 5.39 Výstup analytického nástroja Regression

Grafické znázornenie riešenia príkladu ukazuje obr. 5.40.

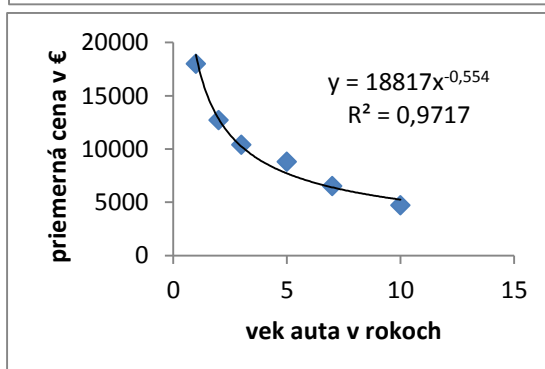
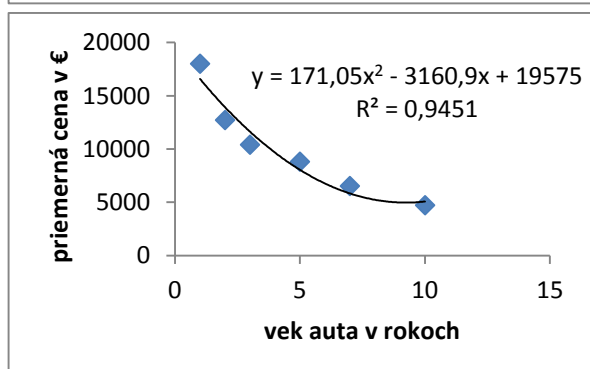
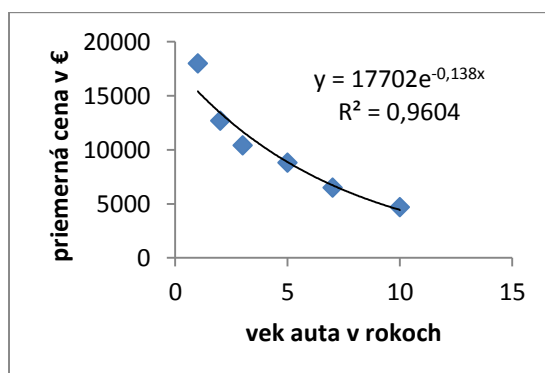
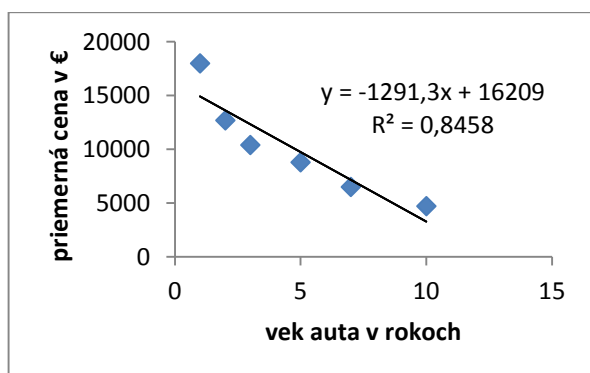


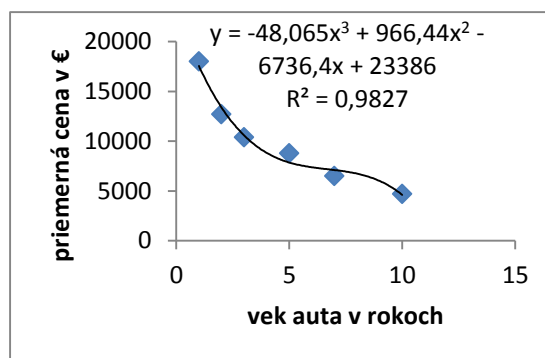
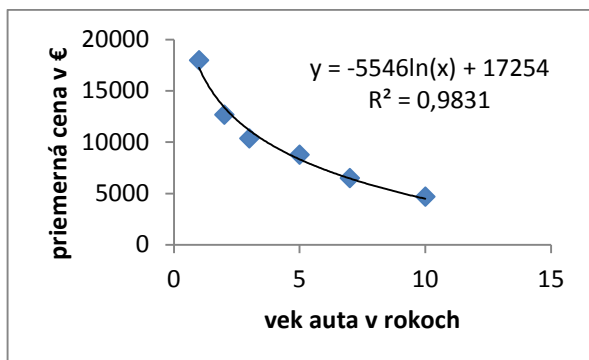
Obr. 5.40 Grafické znázornenie riešenia príkladu

Úlohy

5.8 Porovnajzte, ktorý z nasledujúcich regresných modelov najlepšie popisuje závislosť priemernej ceny vozidiel Škoda Octavia od jeho veku. V rámci riešenia zadanej úlohy boli náhodným výberom namerané nasledujúce priemerné ceny podľa veku vozidiel:

x_i (vek auta v rokoch)	1	2	3	5	7	10
y_i (priemerná cena auta v €)	18000	12700	10400	8800	6500	4700





[Z viacerých regresných modelov je „kvalitnejší“ ten, ktorý má vyšší koeficient determinácie, v našom prípade regresný model popísaný logaritmickou funkciou $y = -5546\ln x + 17254$.]

5.9 Firma má k dispozícii informácie o počte vyrobených výrobkov a spotrebe elektrickej energie za 1 rok.

mesiac	objem výroby (tis. ks)	náklady na elektrickú energiu (tis. €)
január	389,1	145,4
február	379,2	137,3
marec	482,3	162,5
apríl	505	180
máj	574,4	175,4
jún	529	165,2
júl	572,6	176,2
august	641,6	182,4
september	566,8	169
október	538,4	163,9
november	500,6	155,7
december	485,7	154,9

Pomocou nástroja Regresia vyšetrite lineárnu závislosť objemu výroby a nákladov na elektrickú energiu. Určte koeficient korelácie, koeficient determinácie, nájdite rovnicu regresnej priamky na hladine významnosti $\alpha = 0,05$ pomocou celkového F -testu alebo t -testu regresného koeficientu posúďte štatistickú významnosť modelu. Nájdite intervaly spoľahlivosti pre parametre regresnej priamky.

$[r_{xy} = 0,887694; R^2 = 0,788002; y = 80,35545 + 0,162803x; p(F) = p(b_1) = 0,000116 < 0,05 = \alpha \Rightarrow$ model je štatisticky významný, t.j. medzi objemom výroby a nákladmi na elektrickú energiu existuje lineárna závislosť; 95%-ný IS pre $b_0 = \{49,48778; 111,2231\}$ 95%-ný IS pre $b_1 = \{0,103305; 0,222302\}]$

- 5.10** Personálne oddelenie zisťovalo, či existuje vzťah medzi počtom dní absencie v práci a vekom zamestnanca. Náhodne boli vybrané záznamy o dochádzke pätnástich zamestnancov podniku. Pri každom vybranom zamestnancovi bol zaevidovaný údaj o jeho veku a počte dní absencie počas kalendárneho roka. Údaje sú v tabuľke:

vek zamestnanca v rokoch	27	61	37	23	46	58	29	36	64	40	44	25	37	31	30
počet absencií v dňoch	15	6	10	13	9	7	14	11	5	8	10	13	10	12	13

Za predpokladu, že medzi počtom dní absencie a vekom pracovníka je lineárna závislosť, posúďte, či je priama alebo nepriama. Vypočítajte koeficient korelácie a koeficient determinácie.

[nepriama; $r_{xy} = 0,939724$; $R^2 = 0,883082$]

- 5.11** Pri hodnotení skúšok na únavu materiálov možno popísať závislosť počtu kmitov od lomu (Y) na napätí v MPa (X) vhodnou regresnou priamkou. V tabuľke sú uvedené výsledky meraní:

x_i (MPa)	540	560	580	600	630	650	690	700
y_i (počet)	3876	2845	2597	1227	1200	728	654	600

Metódou najmenších štvorcov charakterizujte lineárnu závislosť medzi znakmi X a Y regresnou priamkou $y = b_0 + b_1x$. Vypočítajte koeficient korelácie a index determinácie. Aký počet kmitov môžeme očakávať pri 550 MPa napätí? Akú hodnotu napätia môžeme očakávať pri 1000 kmitov?

[$y = 13548 - 19,122 \cdot x$; $r_{xy} = -0,916554$; $R^2 = 0,840071$; 3030,5; 656,1876 MPa]

- 5.12** Nech X je počet najazdených kilometrov (v km), Y je spotreba pohonných látok (v litroch) náhodne vybraných automobilov. V tabuľke sú uvedené výsledky meraní:

x_i (km)	35	58	64	65	78	96	125	184	228	511
y_i (l)	2,1	3,4	5,6	6,1	7,6	9,2	9,9	15,9	20,1	31,3

Metódou najmenších štvorcov charakterizujte lineárnu závislosť medzi znakmi X a Y regresnou priamkou $y = b_0 + b_1x$. Vypočítajte koeficient korelácie a index determinácie. Akú spotrebu môžeme očakávať pre najazdených 100 km? Akú hodnotu najazdených kilometrov môžeme očakávať pri 17-litrovej spotrebe?

[$y = 2,2599 + 0,0614 \cdot x$; $r_{xy} = 0,946678$; $R^2 = 0,940013$; 8,395696 l; 240,2307 km]

- 5.13** Nech X je vek stroja v rokoch, Y sú týždenné náklady na údržbu a drobné opravy strojov v eurách. V tabuľke sú uvedené výsledky meraní:

x_i (rokov)	1	1	2	3	3	5	6	6
y_i (€)	35	12	75	81	97	100	120	125

Metódou najmenších štvorcov charakterizujte lineárnu závislosť medzi znakmi X a Y regresnou priamkou $y = b_0 + b_1x$. Vypočítajte koeficient korelácie a index determinácie. Akú hodnotu týždenných nákladov môžeme očakávať pre 9-ročný stroj? Akú hodnotu veku stroja môžeme očakávať pri týždenných nákladoch na údržbu 110 eur?

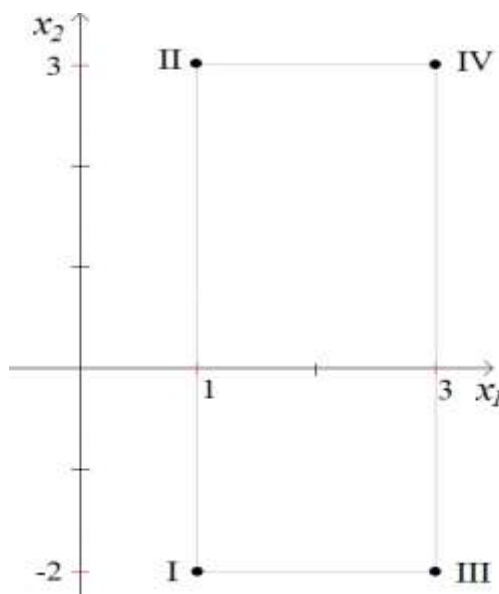
[$y = 21,414 + 17,544 \cdot x$; $r_{xy} = 0,946678$; $R^2 = 0,837003$; 179,3096 €; 5,0493687 r]

5.4 PLÁNOVANIE EXPERIMENTU (DESIGN OF EXPERIMENTS - DOE)

Naplánovať experiment typu n^k znamená určiť n^k rôznych bodov, v ktorých budeme realizovať merania, pričom n je počet úrovní faktorov (počet bodov, ktorými delíme interval na podintervaly rovnakej dĺžky), k je počet faktorov. Počet meraní v každom naplánovanom bode budeme označovať m .

Príklad 5.10 Nech sú dané dva faktory x_1 a x_2 , pričom $x_1 \in \langle 1, 3 \rangle$, $x_2 \in \langle -2, 3 \rangle$. Naplánujme experiment typu 2^2 a typu 3^2 . Nech počet meraní v každom bode bude rovnaký $m = 3$. Koľko získame nameraných hodnôt po vykonaní experimentu?

Riešenie. Máme 2 faktory x_1 a $x_2 \Rightarrow k = 2$. Experiment typu 2^2 znamená, že pre dva faktory x_1 a x_2 budeme merať v koncových bodoch uvedených intervalov (obr.5.41), všetkých bodov bude $2^2 = 4$.



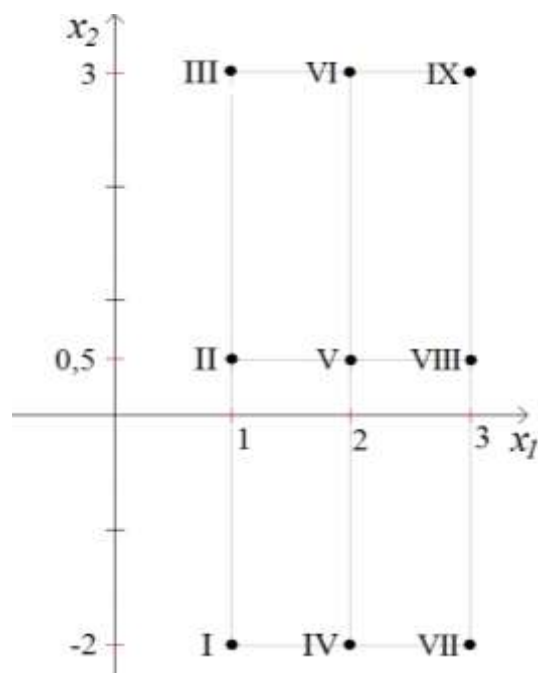
Obr. 5.41 Grafické znázornenie bodov, v ktorých budeme realizovať meranie pri experimente, ktorý je typu 2^2

Experiment typu 2^2 znamená meranie v 4 rôznych bodoch množiny

$$M = \{(1, -2); (1, 3); (3, -2); (3, 3)\}.$$

Celkový počet meraní pre takýto experiment bude $[2^k \cdot m]_{k=2, m=3} = 2^2 \cdot 3 = 12$.

Experiment typu 3^2 znamená, že pre každý faktor x_1 a x_2 berieme hodnoty v koncových bodoch a v strede uvedených intervalov, teda každý z intervalov delíme na $[n - 1]_{n=3} = 3 - 1 = 2$ podintervaly rovnakej dĺžky (obr.5.42). Meranie budeme realizovať v $3^2 = 9$ rôznych bodoch, pre $m = 3$ meraní v každom bode získame $3^2 \cdot 3 = 27$ nameraných hodnôt.



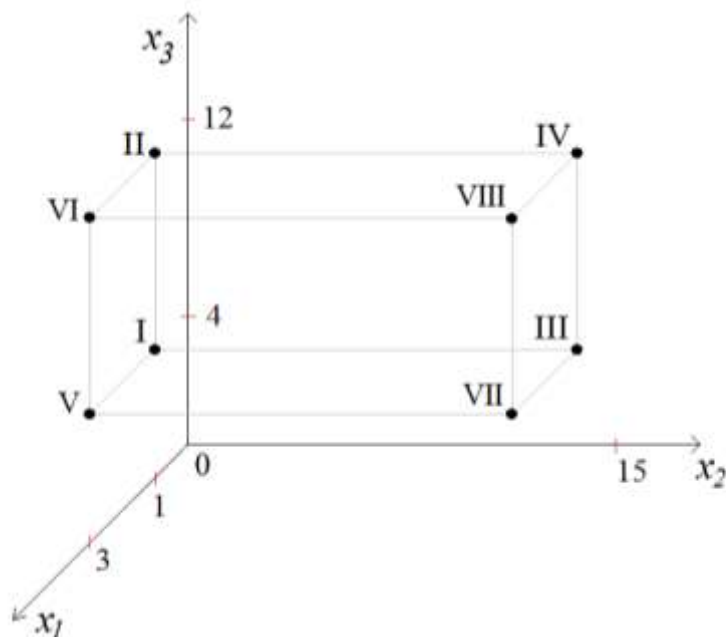
Obr. 5.42 Grafické znázornenie bodov, v ktorých budeme merať pri experimente typu 3^2

Budeme merať 3-krát v každom z bodov množiny

$$M = \{(1; -2); (1; 0,5); (1; 3); (2; -2); (2; 0,5); (2; 3); (3; -2); (3; 0,5); (3; 3)\}.$$

Príklad 5.11 Nech sú dané tri faktory x_1 , x_2 , x_3 , pričom $x_1 \in \langle 1, 3 \rangle$, $x_2 \in \langle 0, 15 \rangle$, $x_3 \in \langle 4, 12 \rangle$. Plánujeme experiment typu 2^3 . V koľkých rôznych bodoch budeme merania realizovať?

Riešenie. Experiment typu 2^3 znamená, že pre tri faktory x_1 , x_2 a x_3 budeme merať v okrajových bodoch uvedených intervalov, takže všetkých bodov bude spolu $2^3 = 8$. Budú to vrcholy kvádra, zobrazeného na obr. 5.43.



Obr. 5.43 Grafické znázornenie bodov, v ktorých budeme realizovať meranie pri experimente, ktorý je typu 2^3

Meranie naplánujeme v bodoch množiny $M = \{(1; 0; 4); (1; 0; 12); (1; 15; 4); (1; 15; 12); (3; 0; 4); (3; 0; 12); (3; 15; 4); (3; 15; 12)\}$.

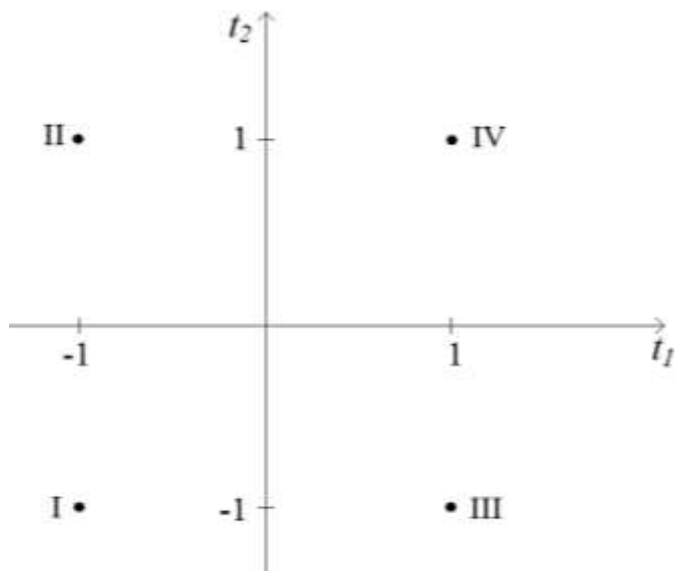
Poznámka 5.7 Uvažujme experiment typu n^k . Ak v každom bode sa robí rovnaký počet meraní, hovoríme, že ide o vyvážený model (experiment). Ak sa nerobí rovnaký počet meraní v každom bode, tak ide o nevyvážený model (experiment).

Veta 5.1 Veta o transformácii – kódovaní. Nech $x \in \langle a, b \rangle$, nech $s = \frac{a+b}{2}$ je stred intervalu $\langle a, b \rangle$, nech $\lambda = \frac{b-a}{2}$ je polovica dĺžky (variačné rozpätie) intervalu $\langle a, b \rangle$. Lineárna transformácia $t(x) = \frac{x-s}{\lambda}$ zobrazí interval $\langle a, b \rangle$ na interval $\langle -1, 1 \rangle$.

$$\langle a, b \rangle \xrightarrow{t(x)} \langle -1, 1 \rangle.$$

Príklad 5.12 Nech sú dané dva faktory x_1 a x_2 , pričom $x_1 \in \langle 1, 3 \rangle$, $x_2 \in \langle -2, 3 \rangle$. Nájdite transformované body experimentu typu 2^2 .

Riešenie. Stred intervalu a variačné rozpätie pre parameter x_1 je $s_1 = \frac{1+3}{2} = 2$, $\lambda_1 = \frac{3-1}{2} = 1$, pre parameter x_2 bude $s_2 = \frac{-2+3}{2} = 0.5$, $\lambda_2 = \frac{3-(-2)}{2} = 2.5$. Odkiaľ $t_1 = \frac{x_1-2}{1}$, $t_2 = \frac{x_2-0.5}{2.5}$. Ľahko sa presvedčíme, že ak $x_1 \in \langle 1, 3 \rangle$, potom $t_1 \in \langle -1, 1 \rangle$ a ak $x_2 \in \langle -2, 3 \rangle$, tak $t_2 \in \langle -1, 1 \rangle$. Preto body meraní transformovaného experimentu typu 2^2 budú postupne $M_t = \{(-1, -1); (-1, 1); (1, -1); (1, 1)\}$, obr. 5.44.



Obr. 5.44 Grafické znázornenie transformovaných bodov pri experimente typu 2^2

Stanovenie regresnej funkcie

Uvažujme experiment typu 2^2 . Označme funkciu $y = f(x_1, x_2) = B_0 + B_1x_1 + B_2x_2 + B_{12}x_1x_2$. Táto funkcia sa nazýva neúplná kvadratická funkcia alebo funkcia lineárneho typu. Koeficienty B_i odhadujeme metódou najmenších štvorcov (MNS). Najskôr však funkciu transformujeme – dostaneme transformovanú (kódovanú) funkciu lineárneho typu $f_t(t_1, t_2) = b_0 + b_1t_1 + b_2t_2 + b_{12}t_1t_2$. MNS tak odhadneme nie koeficienty B_i , ale koeficienty transformovanej funkcie b_i . Odhady koeficientov b_i označme $\hat{b}_i$. Odhad funkcie nazývame regresnou funkciou. Odhad transformovanej funkcie nazývame transformovaná regresná funkcia, ktorá je tiež funkciou lineárneho typu $\hat{y}_T = \hat{f}_t(t_1, t_2) = \hat{b}_0 + \hat{b}_1t_1 + \hat{b}_2t_2 + \hat{b}_{12}t_1t_2$. Na jej základe potom odhadneme pôvodnú regresnú funkciu lineárneho typu $\hat{y} = \hat{f}(x_1, x_2) = \hat{B}_0 + \hat{B}_1x_1 + \hat{B}_2x_2 + \hat{B}_{12}x_1x_2$.

Nech v každom bode robíme m meraní. Celkom máme $2^2 \cdot m$ (všeobecne $2^k \cdot m$) nameraných hodnôt. Pre jednoduchosť nech počet meraní v každom bode sa rovná $m = 1$. Označme vypočítané (vyrovnané, modelované) hodnoty v transformovaných bodoch: $\hat{f}_t(-1, -1) = \hat{y}_1$, $\hat{f}_t(-1, 1) = \hat{y}_2$, $\hat{f}_t(1, -1) = \hat{y}_3$, $\hat{f}_t(1, 1) = \hat{y}_4$.

Odhad $\hat{b}$ -koeficientov transformovanej regresnej funkcie počítame z nameraných y_i hodnôt v transformovaných bodoch:

$$\hat{f}_t(-1, -1) = \hat{y}_1 = \hat{b}_0 + \hat{b}_1 \cdot (-1) + \hat{b}_2 \cdot (-1) + \hat{b}_{12} \cdot (-1) \cdot (-1) = y_1;$$

$$\hat{f}_t(-1, 1) = \hat{y}_2 = \hat{b}_0 + \hat{b}_1 \cdot (-1) + \hat{b}_2 \cdot 1 + \hat{b}_{12} \cdot (-1) \cdot 1 = y_2;$$

$$\hat{f}_t(1, -1) = \hat{y}_3 = \hat{b}_0 + \hat{b}_1 \cdot 1 + \hat{b}_2 \cdot (-1) + \hat{b}_{12} \cdot 1 \cdot (-1) = y_3 ;$$

$$\hat{f}_t(1, 1) = \hat{y}_4 = \hat{b}_0 + \hat{b}_1 \cdot 1 + \hat{b}_2 \cdot 1 + \hat{b}_{12} \cdot 1 \cdot 1 = y_4 .$$

Tejto sústave priradíme matice:

$$\text{matica plánu } X = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \text{ stĺpec meraní } Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \text{ a stĺpec koeficientov } \hat{b} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_{12} \end{pmatrix}.$$

(Vo všeobecnosti X nie je štvorcová matica) . V maticovom tvare $X \cdot \hat{b} = Y$. Túto maticovú rovnicu vynásobíme zľava transponovanou maticou k matici X .

$(X^T \cdot X) \cdot \hat{b} = X^T \cdot Y$; kde X^T je transponovaná matica; $X^T \cdot X$ je štvorcová matica; ak $\exists$ k nej inverzná matica $(X^T \cdot X)^{-1}$ potom

$$(X^T \cdot X)^{-1} \cdot (X^T \cdot X) \cdot \hat{b} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

$$\hat{b} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

Rozpíšme súčin

$$X^T \cdot X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

$$= 4 \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 4 \cdot E = 2^2 E$$

Ak by sme robili m meraní v každom bode, v matici X by sa každý riadok opakoval m – krát, t.j. $X^T \cdot X = m \cdot 2^2 E$. Všeobecne pre experiment 2^k dostaneme $X^T \cdot X = m \cdot 2^k E$. Vzorec $\hat{b} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$ by mal tvar $\hat{b} = (m \cdot 2^k E)^{-1} \cdot X^T \cdot Y$, odkiaľ dostaneme **odhad koeficientov transformovanej regresnej funkcie lineárneho typu**

$$\hat{b} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_{12} \end{pmatrix} = \frac{1}{m \cdot 2^k} X^T Y,$$

kde m je počet meraní, 2^k je typ experimentu, X^T je transponovaná matica plánu, Y je stĺpec meraní.

Skrátená metóda výpočtu

Stĺpec $Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$ meraní nahradíme stĺpcom aritmetických priemerov nameraných hodnôt

a počet meraní považujeme $m = 1$.

Výpočet koeficientov pomocou rozšírenej matice plánu

matica plánu experimentu X

aditívna konštanta	stĺpce samostatného pôsobenia faktorov t_1, t_2	stĺpec spoločného pôsobenia faktorov t_1, t_2	namerané hodnoty y
1	-1	-1	1
1	-1	1	-1
1	1	-1	-1
1	1	1	1

$\sum$

$\begin{pmatrix} 4 & 0 & 0 & 0 \end{pmatrix}$

$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$

Súčet prvkov v 1. stĺpci označíme N .

Súčet prvkov v 2., 3., 4. stĺpci je rovný nule.

Vzorec pre odhad (výpočet) koeficientov: $\hat{b}_i = \frac{1}{N} \sum_j t_{ji} y_j$.

Príklad 5.13 Sleduje sa kvalita používaných pružín na základe toho, koľko "stlačení" (Y) vydržia až do zničenía, v závislosti od dĺžky pružiny ($x_1 \in \{10 \text{ cm}, 15 \text{ cm}\}$) a od hrúbky použitého drôtu ($x_2 \in \{5 \text{ mm}, 7 \text{ mm}\}$).

1. Naplánujte experiment typu 2^2 a určte lineárne transformácie, ktoré zobrazia intervaly $\{10, 15\}$ a $\{5, 7\}$ na interval $\{-1, 1\}$.

2. Metódou najmenších štvorcov odhadnite koeficienty transformovanej regresnej funkcie lineárneho typu, ak v naplánovaných bodoch boli namerané nasledovné hodnoty:

$$y_1 = f(10, 5) = 77; \quad y_3 = f(10, 7) = 98; \quad y_5 = f(15, 5) = 76; \quad y_7 = f(15, 7) = 90;$$

$$y_2 = f(10, 5) = 63; \quad y_4 = f(10, 7) = 82; \quad y_6 = f(15, 5) = 72; \quad y_8 = f(15, 7) = 92.$$

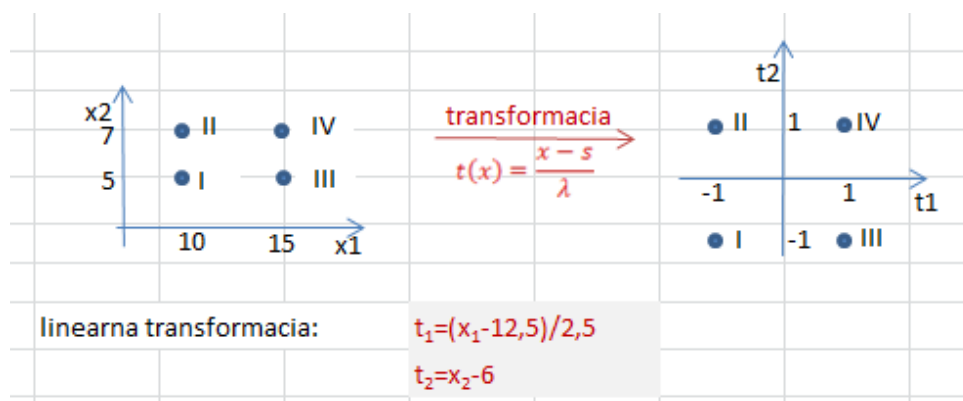
Riešenie. Príklad možno vyriešiť v Exceli využitím funkcie násobenia matic MMULT, funkcie TRANSPOSE, ktorou získame transponovanú maticu X^T k transformovanej matici plánu experimentu X a využitím nami definovaných funkcií. Celé riešenie príkladu je zobrazené v Exceli na nasledujúcich obr. 5.45.

	A	B	C	D	E	F	G
10							
11	x_1 :	10	15	stred=	= (B11+C11)/2	lambda=	= (C11-B11)/2
12	x_2 :	5	7		= (B12+C12)/2		= (C12-B12)/2

	A	B	C	D	E	F	G
10							
11	x_1 :	10	15	stred=	12,5	lambda=	2,5
12	x_2 :	5	7		6		1

Obr. 5.45 Výpočet stredy a variačného rozpätia intervalov faktorov x_1 a x_2

Vzťah, ktorý popisuje použité transformačné vzťahy a transformácia naplánovaných bodov I až IV do nových súradníc t_1, t_2 , sú na obr.5.46.



Obr. 5.46 Grafické zobrazenie bodov I, II, III a IV pred a po transformácii.

Pred začatím výpočtu odhadov koeficientov $\hat{b}_i$ je potrebné napísať transformovanú maticu plánu experimentu X (pole buniek B32:E39), ktorú vytvoríme opísaním koeficientov t_{ij} , ktoré sa nachádzajú pri jednotlivých premenných b_{ij} .

	A	B	C	D	E	F	G	H	I	J	K	L	M
25	MNS odhadneme koeficienty transformovanej regresnej funkcie lineareho typu								$\hat{y}_T = \hat{f}_T(t_1, t_2) = \hat{b}_0 + \hat{b}_1 t_1 + \hat{b}_2 t_2 + \hat{b}_{12} t_1 t_2$				
26													
27		m= 2							I: $f(-1,-1)=$	$b_0+b_1*(-1)+b_2*(-1)+b_{12}*(-1)*(-1)$			
28		2^k= 4							II: $f(-1,1)=$	$b_0+b_1*(-1)+b_2*(1)+b_{12}*(-1)*(1)$			
29		m.2^k= 8							III: $f(1,-1)=$	$b_0+b_1*(1)+b_2*(-1)+b_{12}*(1)*(-1)$			
30									IV: $f(1,1)=$	$b_0+b_1*(1)+b_2*(1)+b_{12}*(1)*(1)$			
31			b_0	b_1	b_2	b_{12}							
32	X=	1	-1	-1	1		Y=	77					
33		1	-1	-1	1			63					
34		1	-1	1	-1			98					
35		1	-1	1	-1			82					
36		1	1	-1	-1			76					
37		1	1	-1	-1			72					
38		1	1	1	1			90					
39		1	1	1	1			92					
40													
41	XT=	=TRANPOSE(B32:E39)											
42													
43													
44													
45													
46	XT*Y=	=MMULT(B41:I44;G32:G39)				b=	=B46/8	b ₀ =	=B46/8				
47							=B47/8	b ₁ =	=B47/8				
48							=B48/8	b ₂ =	=B48/8				
49							=B49/8	b ₁₂ =	=B49/8				
50	odhad transformovanej regresnej funkcie lineareho typu												
51	ft(t ₁ ,t ₂)=	=H46	+	=H47	*t ₁	+	=H48	*t ₂	+	=H49	*t ₁ *t ₂		

	A	B	C	D	E	F	G	H	I	J	K	L	M
25	MNS odhadneme koeficienty transformovanej regresnej funkcie lineareho typu								$\hat{y}_T = \hat{f}_T(t_1, t_2) = \hat{b}_0 + \hat{b}_1 t_1 + \hat{b}_2 t_2 + \hat{b}_{12} t_1 t_2$				
26													
27		m= 2							I: $f(-1,-1)=$	$b_0+b_1*(-1)+b_2*(-1)+b_{12}*(-1)*(-1)$			
28		2^k= 4							II: $f(-1,1)=$	$b_0+b_1*(-1)+b_2*(1)+b_{12}*(-1)*(1)$			
29		m.2^k= 8							III: $f(1,-1)=$	$b_0+b_1*(1)+b_2*(-1)+b_{12}*(1)*(-1)$			
30									IV: $f(1,1)=$	$b_0+b_1*(1)+b_2*(1)+b_{12}*(1)*(1)$			
31			b_0	b_1	b_2	b_{12}							
32	X=	1	-1	-1	1		Y=	77					
33		1	-1	-1	1			63					
34		1	-1	1	-1			98					
35		1	-1	1	-1			82					
36		1	1	-1	-1			76					
37		1	1	-1	-1			72					
38		1	1	1	1			90					
39		1	1	1	1			92					
40													
41	XT=	1	1	1	1	1	1	1	1				
42		-1	-1	-1	-1	1	1	1	1				
43		-1	-1	1	1	-1	-1	1	1				
44		1	1	-1	-1	-1	-1	1	1				
45													
46	XT*Y=	650				b=	81,25	b ₀ =	81,25				
47		10					1,25	b ₁ =	1,25				
48		74					9,25	b ₂ =	9,25				
49		-6					-0,75	b ₁₂ =	-0,75				
50	odhad transformovanej regresnej funkcie lineareho typu												
51	ft(t ₁ ,t ₂)=	81,25	+	1,25	*t ₁	+	9,25	*t ₂	+	-0,75	*t ₁ *t ₂		

Obr. 5.47 Riešenie príkladu v Exceli

Bodový odhad koeficientov $\hat{b}_i$ môžeme v Exceli nájsť aj pomocou rozšírenej matice plánu. Výpočet je zobrazený na nasledujúcom obr.

	G	H	I	J	K	L	M	N	O	P
13		b_0	b_1	b_2	b_{12}					
14		1	t_1	t_2	t_1*t_2	y_i	$1*y_i$	t_1*y_i	t_2*y_i	$t_1*t_2*y_i$
15		1	-1	-1	=I15*J15	77	=H15*\$L15	=I15*\$L15	=J15*\$L15	=K15*\$L15
16		1	-1	-1	=I16*J16	63	=H16*\$L16	=I16*\$L16	=J16*\$L16	=K16*\$L16
17		1	-1	1	=I17*J17	98	=H17*\$L17	=I17*\$L17	=J17*\$L17	=K17*\$L17
18		1	-1	1	=I18*J18	82	=H18*\$L18	=I18*\$L18	=J18*\$L18	=K18*\$L18
19		1	1	-1	=I19*J19	76	=H19*\$L19	=I19*\$L19	=J19*\$L19	=K19*\$L19
20		1	1	-1	=I20*J20	72	=H20*\$L20	=I20*\$L20	=J20*\$L20	=K20*\$L20
21		1	1	1	=I21*J21	90	=H21*\$L21	=I21*\$L21	=J21*\$L21	=K21*\$L21
22		1	1	1	=I22*J22	92	=H22*\$L22	=I22*\$L22	=J22*\$L22	=K22*\$L22
23	Σ :	=SUM(H15:H22)	=SUM(I15:I22)	=SUM(J15:J22)	=SUM(K15:K22)		=SUM(M15:M22)	=SUM(N15:N22)	=SUM(O15:O22)	=SUM(P15:P22)
24							$b_0=$	$b_1=$	$b_2=$	$b_{12}=$
25							=M23/\$H\$23	=N23/\$H\$23	=O23/\$H\$23	=P23/\$H\$23

	b_0	b_1	b_2	b_{12}					
	1	t_1	t_2	t_1*t_2	y_i	$1*y_i$	t_1*y_i	t_2*y_i	$t_1*t_2*y_i$
	1	-1	-1	1	77	77	-77	-77	77
	1	-1	-1	1	63	63	-63	-63	63
	1	-1	1	-1	98	98	-98	98	-98
	1	-1	1	-1	82	82	-82	82	-82
	1	1	-1	-1	76	76	76	-76	-76
	1	1	-1	-1	72	72	72	-72	-72
	1	1	1	1	90	90	90	90	90
	1	1	1	1	92	92	92	92	92
Σ :	8	0	0	0		650	10	74	-6
						$b_0=$	$b_1=$	$b_2=$	$b_{12}=$
						81,25	1,25	9,25	-0,75

Obr. 5.48 Výpočet koeficientov pomocou rozšírenej matice plánu

Vyhodnotenie experimentu

Z hľadiska vyhodnocovania experimentu uvažujeme:

Ak b_i je v absolútnej hodnote malé číslo (vyjadríme zápisom $b_i = 0$), potom faktor t_i nemá významný vplyv na hodnotu transformovanej funkcie $f_t(t_1, t_2)$, ak b_i je v absolútnej hodnote veľké číslo ($b_i \neq 0$), potom faktor t_i má veľký vplyv na hodnotu transformovanej funkcie $f_t(t_1, t_2)$. Rovnaké úvahy platia aj pre koeficient b_{ij} , ktorým popisujeme veľkosť vzájomného pôsobenia faktorov t_i a t_j .

Tvrdenia $b_i = 0$, resp. $b_i \neq 0$ sú pre nás hypotézami, ktorých platnosť musíme testovať. V praxi hodnoty b -koeficientov nepoznáme, poznáme iba ich odhady $\hat{b}$, ktoré použijeme pri výpočte testovacej štatistiky.

1. Štatistická významnosť jednotlivých parametrov (faktorov) lineárneho modelu pre experiment typu 2^2

Individuálny t-test – štatistická významnosť vzťahu Y (závislej premennej) s určitou nezávislou premennou x_i v danom modeli. Určuje sa hladina významnosti α , t.j. pravdepodobnosť, že zamietame nulovú hypotézu, ak je pravdivá. Určuje sa hypotéza H_0 : Regresný koeficient nie je štatisticky významný ($b_i = 0, i = 1, 2, 12$), (faktor t_i nemá veľký vplyv na funkciu f_t , t.j. x_i nemá veľký vplyv na funkciu f – vplyv nie je štatisticky významný)

H_1 : Regresný koeficient je štatisticky významný ($b_i \neq 0, i = 1, 2, 12$), (faktor t_i má veľký vplyv na funkciu f_t , t.j. x_i má veľký vplyv na funkciu f – vplyv je štatisticky významný).

Krátko zapísané: $H_0: b_i = 0 >< H_1: b_i \neq 0$ pre $i = 1, 2, 12$.

Testovacia charakteristika (štatistika):

$$t = \frac{\hat{b} \cdot 2^k \cdot \sqrt{m(m-1)}}{R_0},$$

kde 2^k je typ experimentu, m je počet meraní, R_0^2 je súčet reziduálnych štvorcov

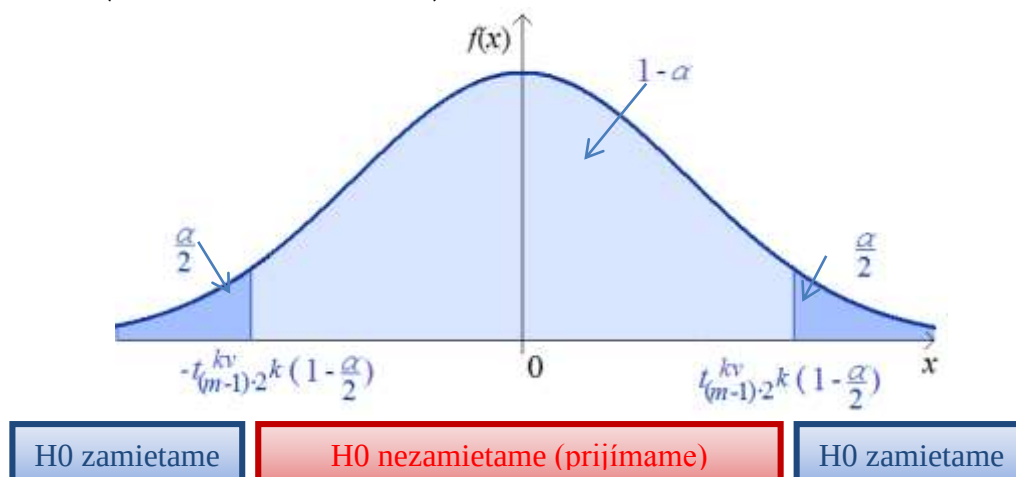
$$R_0^2 = \sum_{i=1}^{m \cdot 2^k} (y_i - \hat{y}_i)^2,$$

kde y_i sú namerané hodnoty, $\hat{y}_i$ sú vyrovnané teoretické hodnoty.

Testovacia charakteristika t má Studentovo t -rozdelenie so stupňom voľnosti $(m-1) \cdot 2^k$.

O výsledku testu rozhodujeme pomocou:

a) kritickej oblasti (kritického oboru) W : Ak $|t| > t_{(m-1) \cdot 2^k}^{kv} \left(1 - \frac{\alpha}{2}\right)$ alebo ak $t_{(m-1) \cdot 2^k}^{kv} \left(1 - \frac{\alpha}{2}\right) < t < t_{(m-1) \cdot 2^k}^{kv} \left(1 - \frac{\alpha}{2}\right)$ patrí $W = \left(-\infty; -t_{(m-1) \cdot 2^k}^{kv} \left(1 - \frac{\alpha}{2}\right)\right) \cup \left(t_{(m-1) \cdot 2^k}^{kv} \left(1 - \frac{\alpha}{2}\right); \infty\right)$ zamietame H_0 .



b) intervalu spoľahlivosti pre b - koeficienty:

V prípade, že číslo 0 ($H_0: b_i = 0$) leží v $(1 - \alpha)$ %-nom intervale spoľahlivosti, na hladine významnosti α prijímame H_0 .

Riešenie v Exceli: Data→Data analysis→Regression – automaticky hľadá 95%-ný interval spoľahlivosti (možno zadať aj inú spoľahlivosť).

c) p -hodnoty: ak je $p \leq \alpha$ zamietame H_0 , ak platí $p > \alpha$ nezamietame H_0 ("považuje sa za pravdivú" - čím menšia je p -hodnota, tým viac sme presvedčení, že H_0 nie je správna a mala by byť prijatá alternatívna hypotéza H_1).

V Exceli volíme postupnosť krokov: Data→Data analysis→Regression.

2. Štatistická významnosť regresného modelu

Celkový F -test – významnosť regresného vzťahu Y (závislej premennej) a x_1, x_2 (nezávislých premenných). Určuje (volí) sa hladina významnosti α . Testuje sa hypotéza H_0 : Regresný model nie je štatisticky významný (t.j. všetky regresné koeficienty $\hat{b}_i = 0, i = 1, 2, 12$) proti hypotéze.

H_1 : Regresný model je štatisticky významný (aspoň jeden regresný koeficient $\hat{b}_i \neq 0$).

Krátko zapísané $H_0: \hat{b}_1 = \hat{b}_2 = \hat{b}_{12} = 0 > < H_1: \exists$ aspoň jedno $i: \hat{b}_i \neq 0$.

Zamietnutie H_0 znamená, že aspoň jedna z premenných vplyva významne na Y .

(v Exceli: Data→Data analysis →Regression→ANOVA)

Testovacia charakteristika (štatistika):

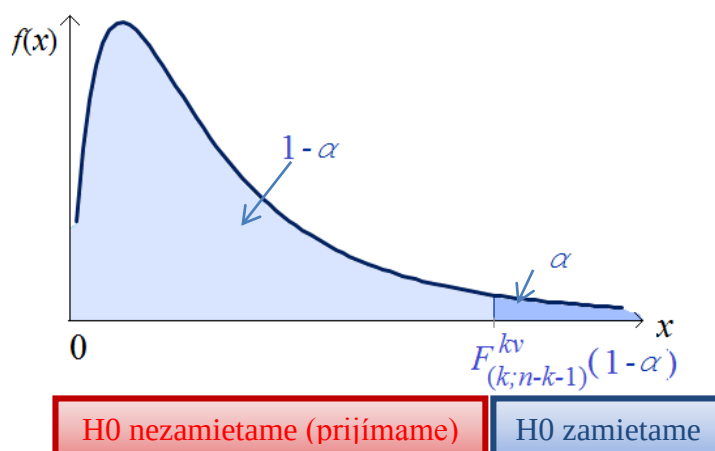
$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{(n-k-1) \cdot \sum_{i=1}^{m \cdot 2^k} (\hat{y}_i - \bar{y})^2}{k \sum_{i=1}^{m \cdot 2^k} (y_i - \hat{y}_i)^2},$$

kde n je rozsah výberového súboru ($n = m \cdot 2^k$), k je počet faktorov, $\bar{y}$ je priemer nameraných hodnôt, y_i je i -tá nameraná hodnota, $\hat{y}_i$ je i -tá modelovaná hodnota.

Testovacia štatistika F má Fisherovo rozdelenie so stupňami voľnosti $k, n - k - 1$.

Rozhodovanie o výsledku testu pomocou:

a) kritického oboru : Ak $F > F_{(k,n-k-1)}^{kv}(1 - \alpha)$, H_0 zamietame;



b) p -hodnoty: ak $p \leq \alpha$, zamietame H_0 (ANOVA v Exceli).

Príklad 5.14 Sledovali sme opotrebovanie pneumatiky pomocou hĺbky dezénu pneumatiky v mm v závislosti od faktorov x_1, x_2 , kde x_1 je počet najazdených km v tisíc km ($x_1 \in \langle 10, 70 \rangle$) a x_2 je vek pneumatiky v rokoch ($x_2 \in \langle 5, 10 \rangle$). Uvažujme experiment typu 2^2 a regresnú funkciu $f(x_1, x_2)$ lineárneho typu, pričom v každom zo 4 bodov sme vykonali po 3 merania, čím sme získali nasledujúce hodnoty:

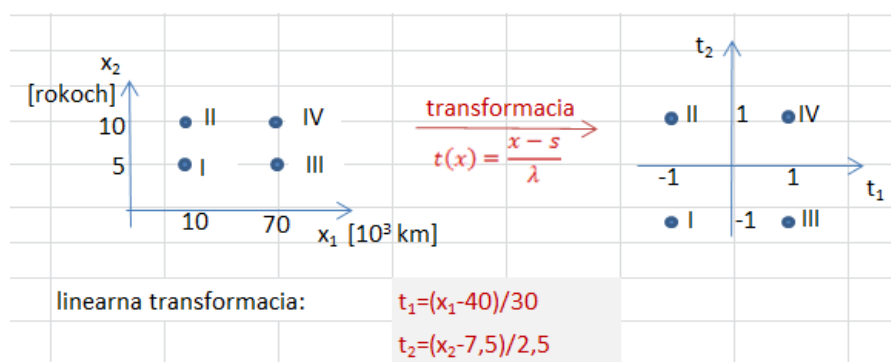
$$\begin{aligned} y_1 = f(10, 5) &= 3; & y_4 = f(10, 10) &= 2,4; & y_7 = f(70, 5) &= 2,25; & y_{10} = f(70, 10) &= 1,95; \\ y_2 = f(10, 5) &= 3,27; & y_5 = f(10, 10) &= 3,09; & y_8 = f(70, 5) &= 2,01; & y_{11} = f(70, 10) &= 1,5; \\ y_3 = f(10, 5) &= 3,5; & y_6 = f(10, 10) &= 3,025; & y_9 = f(70, 5) &= 2; & y_{12} = f(70, 10) &= 1,7. \end{aligned}$$

1. Metódou najmenších štvorcov odhadnite koeficienty regresnej funkcie lineárneho typu.
2. Na hladine významnosti 5% otestujte, ktoré faktory sú rozhodujúce pre hĺbku dezénu pneumatiky.
3. Pomocou koeficientu (indexu) determinácie charakterizujte celkovú štatistickú významnosť daného modelu lineárneho typu.
4. Na hladine významnosti 5% pomocou F -testu otestujte celkovú štatistickú významnosť (signifikanciu) daného modelu lineárneho typu.

Riešenie.

Namerané hodnoty v transformovaných bodoch sú:

$$\begin{aligned} y_1 = f_t(-1, -1) &= 3; & y_4 = f_t(-1, 1) &= 2,4; & y_7 = f_t(1, -1) &= 2,25; & y_{10} = f_t(1, 1) &= 1,95; \\ y_2 = f_t(-1, -1) &= 3,27; & y_5 = f_t(-1, 1) &= 3,09; & y_8 = f_t(1, -1) &= 2,01; & y_{11} = f_t(1, 1) &= 1,5; \\ y_3 = f_t(-1, -1) &= 3,5; & y_6 = f_t(-1, 1) &= 3,025; & y_9 = f_t(1, -1) &= 2; & y_{12} = f_t(1, 1) &= 1,7. \end{aligned}$$



Obr. 5.49 Grafické zobrazenie bodov I, II, III a IV pred a po transformácii.

1. Transformovaná regresná funkcia lineárneho typu $f_t(t_1, t_2)$ má tvar $f_t(t_1, t_2) = b_0 + b_1 t_1 + b_2 t_2 + b_{12} t_1 t_2$.

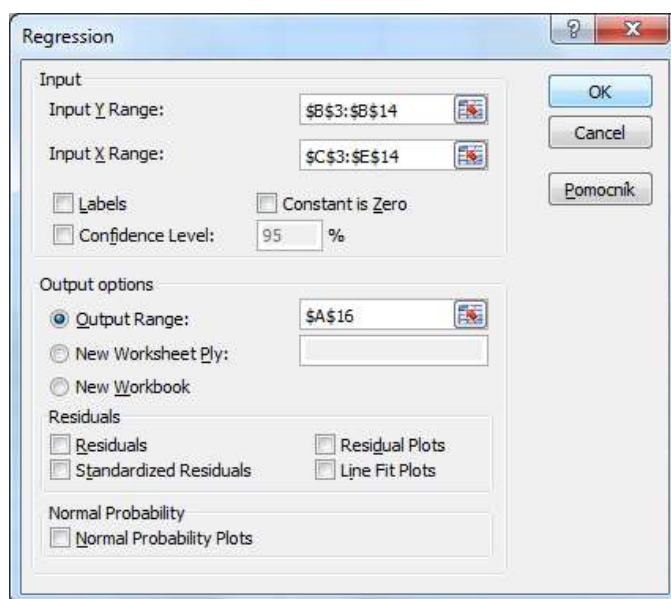
Na výpočet odhadu koeficientov transformovanej regresnej funkcie lineárneho typu $f_t(t_1, t_2)$ použijeme analytický nástroj Regression. Pred začatím výpočtu je potrebné zadať vstupné parametre v tvare vektora závislej premennej Y – nameraných hodnôt (pole B3:B14) a v tvare poľa nezávislých premenných (pole C3:E14) – v našom prípade vektora samostatného pôsobenia transformovaného faktora t_1 , vektora samostatného pôsobenia transformovaného

faktora t_2 a vektora popisujúceho spoločné pôsobenie transformovaných faktorov t_1 a t_2 , obr. 5.50.

	A	B	C	D	E
1					
2		y_i	t_1	t_2	$t_1 * t_2$
3	I:	3	-1	-1	1
4		3,27	-1	-1	1
5		3,5	-1	-1	1
6	II:	2,4	-1	1	-1
7		3,09	-1	1	-1
8		3,025	-1	1	-1
9	III:	2,25	1	-1	-1
10		2,01	1	-1	-1
11		2	1	-1	-1
12	IV:	1,95	1	1	1
13		1,5	1	1	1
14		1,7	1	1	1

Obr. 5.50 Syntax zápisu vstupných údajov pri použití analytického nástroja Regression

Umiestnenie vstupných údajov je potrebné zadať do dialógového okna, obr.5.51.



Obr. 5.51 Dialógové okno analytického nástroja Regression

Štandardizovaným výstupom nástroja regresnej analýzy je tabuľka na obr. 5.52.

Regression Statistics								
Multiple R	0,942193							
R Square	0,8877277							
Adjusted R Square	0,8456256							
Standard Error	0,2639405							
Observations	12							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	3	4,40665625	1,468885	21,0851	0,000373021			
Residual	8	0,557316667	0,069665					
Total	11	4,963972917						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
Intercept	2,4745833	0,076193057	32,47781	8,8E-10	2,298881829	2,6502848	2,29888183	2,650284838
X Variable 1	-0,5729167	0,076193057	-7,519277	6,8E-05	-0,748618171	-0,397215	-0,7486182	-0,397215162
X Variable 2	-0,1970833	0,076193057	-2,586631	0,03228	-0,372784838	-0,021382	-0,3727848	-0,021381829
X Variable 3	0,0120833	0,076193057	0,158588	0,87792	-0,163618171	0,1877848	-0,1636182	0,187784838

Obr. 5.52 Výstup analytického nástroja Regression

V tabuľke stĺpec Coefficients obsahuje bodový odhad $\hat{b}$ koeficientov transformovanej regresnej funkcie lineárneho typu $f_t(t_1, t_2)$: $\hat{b}_0 = 2,4745833$, $\hat{b}_1 = -0,5729167$, $\hat{b}_2 = -0,1970833$, $\hat{b}_{12} = 0,0120833$.

Bodový odhad transformovanej regresnej funkcie lineárneho typu

$$\hat{f}_t(t_1, t_2) = 2,4745833 - 0,5729167 \cdot t_1 - 0,1970833 \cdot t_2 + 0,0120833 \cdot t_1 \cdot t_2.$$

Po spätnej transformácii premenných $t_1 = \frac{x_1 - 40}{30}$ a $t_2 = \frac{x_2 - 7,5}{2,5}$ získame odhad pôvodnej regresnej funkcie lineárneho typu

$$\hat{f}(x_1, x_2) = \hat{f}\left(\frac{x_1 - 40}{30}, \frac{x_2 - 7,5}{2,5}\right) = 3,87805556 - 0,0203056 \cdot x_1 - 0,0852778 \cdot x_2 + 0,00016111 \cdot x_1 \cdot x_2.$$

2. Na hladine významnosti 5% otestujeme, ktoré faktory sú rozhodujúce pre hĺbku dezénu pneumatiky.

Vyhodnotenie vplyvu jednotlivých faktorov na hĺbku dezénu pneumatiky urobíme pomocou testu hypotéz. Samostatný vplyv faktora t_1 na hodnotu $f_t(t_1, t_2)$ testujeme pomocou nulovej hypotézy $H_0: b_1 = 0$ oproti alternatívnej $H_1: b_1 \neq 0$.

a) Rozhodovanie o výsledku testu pomocou kritického oboru.

Hodnotu testovacieho kritéria prečítame zo stĺpca t Stat v riadku koeficientu b_1 . Pre určenie kritickej oblasti W nájdeme kvantil t -rozdelenia $t_{(m-1) \cdot 2^k}^{kv} \left(1 - \frac{\alpha}{2}\right) = t_{(3-1) \cdot 2^2}^{kv} \left(1 - \frac{0,05}{2}\right) = t_8^{kv}(0,975) = T.INV(0,975;8) = 2,306004135$. Kritická oblasť W , oblasť zamietnutia nulovej hypotézy je $W = (-\infty; -2,306004135) \cup (2,306004135, \infty)$.

Hodnotu testovacieho kritéria $t = -7,519277$ patrí do kritickej oblasti W , čo znamená, že nulovú hypotézu zamietame a nezamietame alternatívnu hypotézu.

b) Rozhodovanie o výsledku testu pomocou p -hodnoty.

P -hodnotu (p -value) prečítame zo stĺpca P -value v riadku koeficientu b_1 a porovnáваме s hladinou významnosti α . Hodnota $p = 6,8E - 05$ je menšia ako zvolená hladina významnosti α , preto zamietame nulovú hypotézu $H_0: b_1 = 0$.

c) Rozhodovanie o výsledku testu pomocou intervalu spoľahlivosti.

Dolnú a hornú hranicu 95%-ného intervalu spoľahlivosti pre koeficient b_1 prečítame z tabuľky v stĺpcoch Lower 95% a Upper 95% v riadku koeficienta b_1 . Číslo nula neleží v intervale spoľahlivosti $\{-0,748618171; -0,3972152\}$, preto zamietame nulovú hypotézu $H_0: b_1 = 0$.

Záver: Na hladine významnosti 5% možno konštatovať, že samostatné pôsobenie faktora t_1 má na hodnotu transformovanej funkcie $f_t(t_1, t_2)$ podstatný vplyv. Slovné môžeme sformulovať, že bol štatisticky preukázaný nezanedbateľný vplyv faktora x_1 (počtu najazdených km) na hodnotu regresnej funkcie $f(x_1, x_2)$ (na hĺbku dezénu pneumatiky).

Samostatný vplyv faktora t_2 na hodnotu $f_t(t_1, t_2)$ testujeme pomocou nulovej hypotézy $H_0: b_2 = 0$ oproti alternatívnej $H_1: b_2 \neq 0$.

Rozhodovanie o výsledku testu pomocou kritického oboru.

Hodnotu testovacieho kritéria prečítame zo stĺpca t Stat v riadku koeficientu b_2 . Pre určenie kritickej oblasti W nájdeme kvantil t -rozdelenia $t_{(m-1) \cdot 2^k}^{kv} \left(1 - \frac{\alpha}{2}\right) = t_{(3-1) \cdot 2^2}^{kv} \left(1 - \frac{0,05}{2}\right) = t_8^{kv}(0,975) = T.INV(0,975;8) = 2,306004135$. Kritická oblasť W , oblasť zamietnutia nulovej hypotézy je $W = (-\infty; -2,306004135) \cup (2,306004135, \infty)$.

Hodnota testovacieho kritéria $t = -2,586631$ patrí do kritickej oblasti W , čo znamená, že nulovú hypotézu zamietame a nezamietame alternatívnu hypotézu.

Rozhodovanie o výsledku testu pomocou p -hodnoty.

P -hodnotu (p -value) prečítame zo stĺpca P -value v riadku koeficientu b_2 a porovnáваме s hladinou významnosti α . Hodnota $p = 0,032282$ je menšia ako zvolená hladina významnosti α , preto zamietame nulovú hypotézu $H_0: b_2 = 0$.

Rozhodovanie o výsledku testu pomocou intervalu spoľahlivosti.

Dolnú a hornú hranicu 95%-ného intervalu spoľahlivosti pre koeficient b_2 prečítame z tabuľky v stĺpcoch Lower 95% a Upper 95% v riadku koeficienta b_2 . Číslo nula neleží v intervale spoľahlivosti $\{-0,372784838; -0,0213818\}$, preto zamietame nulovú hypotézu $H_0: b_2 = 0$.

Záver: Na hladine významnosti 5% možno konštatovať, že samostatné pôsobenie faktora t_2 má na hodnotu transformovanej funkcie $f_t(t_1, t_2)$ štatisticky významný vplyv. Slovné môžeme

sformulovať, že bol štatisticky preukázaný zanedbateľný vplyv faktora x_2 (vek pneumatiky) na hodnotu regresnej funkcie $f(x_1, x_2)$ (na hĺbku dezénu pneumatiky).

Vplyv spoločného pôsobenia faktorov t_1, t_2 na hodnotu $f_t(t_1, t_2)$ otestujeme pomocou nulovej hypotézy $H_0: b_{12} = 0$ oproti alternatívnej $H_1: b_{12} \neq 0$.

Rozhodovanie o výsledku testu pomocou kritického oboru.

Hodnotu testovacieho kritéria prečítame zo stĺpca t Stat v riadku koeficientu b_{12} . Pre určenie kritickej oblasti W nájdeme kvantil t -rozdelenia $t_{(m-1) \cdot 2^k}^{kv} \left(1 - \frac{\alpha}{2}\right) = t_{(3-1) \cdot 2^2}^{kv} \left(1 - \frac{0,05}{2}\right) = t_8^{kv}(0,975) = \text{T.INV}(0,975;8) = 2,306004135$. Kritická oblasť W , oblasť zamietnutia nulovej hypotézy je $W = (-\infty; -2,306004135) \cup (2,306004135, \infty)$.

Hodnota testovacieho kritéria $t = 0,1585884$ nepatrí do kritickej oblasti W , čo znamená, že nulovú hypotézu nezamietame a zamietame alternatívnu hypotézu.

Rozhodovanie o výsledku testu pomocou p -hodnoty.

P -hodnotu (p -value) prečítame zo stĺpca P -value v riadku koeficientu b_{12} a porovnávame s hladinou významnosti α . Hodnota $p = 0,877923$ je väčšia ako zvolená hladina významnosti α , preto nezamietame nulovú hypotézu $H_0: b_{12} = 0$.

Rozhodovanie o výsledku testu pomocou intervalu spoľahlivosti.

Dolnú a hornú hranicu 95%-ného intervalu spoľahlivosti pre koeficient b_{12} prečítame z tabuľky v stĺpcoch Lower 95% a Upper 95% v riadku koeficienta b_{12} . Číslo nula leží v intervale spoľahlivosti $\langle -0,163618171; 0,18778484 \rangle$, preto nezamietame nulovú hypotézu $H_0: b_{12} = 0$.

Záver: Na hladine významnosti 5% možno konštatovať, že spoločné pôsobenie faktorov t_1, t_2 nemá na hodnotu transformovanej funkcie $f_t(t_1, t_2)$ štatisticky významný vplyv. Slovné môžeme sformulovať, že bol štatisticky preukázaný zanedbateľný vplyv faktorov x_1, x_2 (spoločné pôsobenie počtu najazdených kilometrov a veku pneumatiky) na hodnotu regresnej funkcie $f(x_1, x_2)$ (na hĺbku dezénu pneumatiky).

3. Pomocou koeficientu (indexu) determinácie charakterizujeme celkovú štatistickú významnosť daného modelu lineárneho typu.

Index determinácie sa nachádza v časti tabuľky Regression Statistics v riadku R Square. $R^2 = 0,8877277$, znamená, že 88,77% variability hodnôt premennej Y je vyjadrených regresným modelom.

4. Na hladine významnosti 5% pomocou F -testu otestujeme celkovú štatistickú významnosť regresného vzťahu Y (závislej premennej) a x_1, x_2 (nezávislých premenných) – t.j. celkovú štatistickú významnosť (signifikanciu) daného modelu lineárneho typu.

Významnosť celkového modelu zistíme testovaním hypotéz $H_0: b_1 = b_2 = b_{12} = 0$ oproti alternatívnej H_1 : aspoň jeden z uvedených koeficientov nie je nulový. Hodnotu testovacieho kritéria F prečítame v tabuľky ANOVA v riadku Regression a stĺpci F a porovnávame s α kvantilom Fisherovho F -rozdelenia $F_{(k,n-k-1)}^{kv}(1-\alpha) = F_{(3,12-3-1)}^{kv}(1-0,05) = F_{(3,8)}^{kv}(0,95) = F.INV(0,95; 3; 8) = 4,066181$. $F = 21,08511 > 4,066181 = F_{(3,8)}^{kv}(0,95)$, preto zamietame H_0 a nezamietame H_1 . Rozhodovanie o teste môžeme urobiť aj pomocou p -hodnoty, ktorú prečítame v stĺpci Significance F tabuľky ANOVA. V našom prípade $p = 0,000373021$, čo je menšie ako hladina významnosti $\alpha = 0,05$, preto nulovú hypotézu zamietame. Môžeme konštatovať, že na hladine významnosti $\alpha = 0,05$ regresný model je štatisticky významný (signifikantný).

Úlohy

5.14 Uvažujme experiment typu 2^2 a transformovanú regresnú funkciu lineárneho typu $f_t(t_1, t_2)$, pričom v každom zo 4 bodov sme vykonali po 2 merania, čím sme získali nasledovné hodnoty v transformovaných bodoch:

I	II	III	IV
34	45	36	43
36	47	37	44

Odhadnite MNŠ rovnicu transformovanej regresnej funkcie lineárneho typu $f_t(t_1, t_2)$.

$$[f(t_1, t_2) = 40,25 - 0,25 \cdot t_1 + 4,5 \cdot t_2 - t_1 \cdot t_2]$$

5.15 Uvažujme experiment typu 2^2 a transformovanú regresnú funkciu lineárneho typu $f_t(t_1, t_2)$, pričom v každom zo 4 bodov sme vykonali po 2 merania, čím sme získali nasledovné hodnoty v transformovaných bodoch:

I	II	III	IV
23	27	29	33
25	27	25	34

Odhadnite MNŠ rovnicu transformovanej regresnej funkcie lineárneho typu $f_t(t_1, t_2)$.

Na hladine významnosti $\alpha = 0,03$ otestujte vplyv samostatného pôsobenia faktorov t_1, t_2 a spoločného pôsobenia faktorov t_1, t_2 na hodnotu $f_t(t_1, t_2)$.

$[f(t_1, t_2) = 27,875 + 2,375 \cdot t_1 + 2,375 \cdot t_2 + 0,875 t_1 \cdot t_2;$ vplyv samostatného pôsobenia faktora t_1 a samostatného pôsobenia faktora t_2 sú štatisticky významné, vplyv spoločného pôsobenia faktorov t_1, t_2 nie je štatisticky významný.]

5.16 Fyzikálne vlastnosti radia drevo medzi najlepšie stavebné materiály. Budeme preto skúmať závislosť medzi jeho rigiditou, hustotou a jeho elasticitou. Hľadáme závislosť rigidity od hustoty a elasticity drevín. Na vybraných vzorkách meriame Y – rigiditu dreva v kg/m^2 , x_1 hustotu dreva v kg/m^2 ($x_1 \in \langle 400, 700 \rangle$), x_2 elasticitu dreva v kg/m^2 ($x_2 \in \langle 500, 1100 \rangle$). Plánovaný je experiment typu 2^2 . Počet meraní v každom naplánovanom bode je 4. Namerané hodnoty v transformovaných bodoch sú:

I	II	III	IV
4882	952	864	1162
4881	981	889	1100
4900	964	891	1127
4880	942	815	1098

Nájdite metódou najmenších štvorcov odhad transformovanej regresnej funkcie lineárneho typu. Na hladine významnosti $\alpha = 0,01$ vyhodnoťte experiment.

$[f(t_1, t_2) = 1958 - 964,75 \cdot t_1 - 917,25 \cdot t_2 + 1045,75 t_1 \cdot t_2$. Na hladine významnosti 1% možno konštatovať, že vplyv samostatného pôsobenia faktora t_1 , aj samostatného pôsobenia faktora t_2 , aj spoločného pôsobenia faktorov t_1, t_2 na hodnotu transformovanej funkcie $f_t(t_1, t_2)$ je signifikantný. Bol štatisticky preukázaný nezanedbateľný samostatný vplyv hustoty dreva, samostatný vplyv elasticity dreva aj ich spoločné pôsobenie na hodnotu regresnej funkcie lineárneho typu $f(x_1, x_2)$ (na rigiditu dreva). Na hladine významnosti $\alpha = 0,01$ možno konštatovať, že regresný model lineárneho typu je štatisticky významný.]

PRÍLOHY

PRÍLOHA 1

Aplikácia symbolu súčtu

Príklad 1

Rozpíšte výrazy:

- a) $\sum_{i=1}^5 x_i$; b) $\sum_{i=1}^n x_i$; c) $\sum_{i=1}^3 x_i^2$; d) $\sum_{i=1}^n x_i^4$;
e) $\sum_{i=1}^k x_i n_i$; f) $\sum_{i=1}^k n$; g) $\sum_{i=1}^n 1$;

Riešenie.

- a) $\sum_{i=1}^5 x_i = x_1 + x_2 + x_3 + x_4 + x_5$;
b) $\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n$;
c) $\sum_{i=1}^3 x_i^2 = x_1^2 + x_2^2 + x_3^2$;
d) $\sum_{i=1}^n x_i^4 = x_1^4 + x_2^4 + x_3^4 + \dots + x_n^4$;
e) $\sum_{i=1}^k x_i n_i = x_1 n_1 + x_2 n_2 + x_3 n_3 + \dots + x_k n_k$;
f) $\sum_{i=1}^k n = n + n + n + \dots + n = k \cdot n$;
g) $\sum_{i=1}^n 1 = 1 + 1 + 1 + \dots + 1 = n \cdot 1 = n$;

Príklad 2

Rozpíšte výrazy:

- a) $\sum_{i=1}^n (x_i - 3)^2$; b) $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$; c) $\frac{\sum_{i=1}^l x_i^k n_i}{\sum_{i=1}^l n_i}$;
d) $\sum_{i=1}^n c \frac{1}{x_i^2}$; e) $\sqrt{\sum_{i=1}^n x_i^2}$; f) $\sum_{i=1}^n \sqrt{x_i}$;

Riešenie.

- a) $\sum_{i=1}^n (x_i - 3)^2 = (x_1 - 3)^2 + (x_2 - 3)^2 + (x_3 - 3)^2 + \dots + (x_n - 3)^2$;
b) $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$;
c) $\frac{\sum_{i=1}^l x_i^k n_i}{\sum_{i=1}^l n_i} = \frac{x_1^k n_1 + x_2^k n_2 + x_3^k n_3 + \dots + x_l^k n_l}{n_1 + n_2 + n_3 + \dots + n_l}$;
d) $\sum_{i=1}^n c \frac{1}{x_i^2} = c \frac{1}{x_1^2} + c \frac{1}{x_2^2} + c \frac{1}{x_3^2} + \dots + c \frac{1}{x_n^2} = c \left(\frac{1}{x_1^2} + \frac{1}{x_2^2} + \frac{1}{x_3^2} + \dots + \frac{1}{x_n^2} \right) = c \sum_{i=1}^n \frac{1}{x_i^2}$;

$$e) \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2};$$

$$f) \sum_{i=1}^n \sqrt{x_i} = \sqrt{x_1} + \sqrt{x_2} + \sqrt{x_3} + \dots + \sqrt{x_n};$$

Príklad 3

Napište v skrátenej tvare pomocou symbolu súčtu:

$$a) x_1 y_1 + x_2 y_2 + x_3 y_3 + \dots + x_n y_n;$$

$$b) \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}};$$

$$c) x_1^2 y_1 + x_2^2 y_2 + x_3^2 y_3 + \dots + x_n^2 y_n;$$

$$d) n \cdot 2 = 2 + 2 + 2 + \dots + 2;$$

$$e) \sqrt{(x_1 - 2)^2 + (x_2 - 2)^2 + (x_3 - 2)^2 + \dots + (x_n - 2)^2};$$

$$f) \frac{y_1^2 + y_2^2 + y_3^2}{3} - \left(\frac{y_1 + y_2 + y_3}{3} \right)^2;$$

Riešenie.

$$a) x_1 y_1 + x_2 y_2 + x_3 y_3 + \dots + x_n y_n = \sum_{i=1}^n x_i y_i;$$

$$b) \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}} = \sum_{i=1}^n \frac{n}{\frac{1}{x_i}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}};$$

$$c) x_1^2 y_1 + x_2^2 y_2 + x_3^2 y_3 + \dots + x_n^2 y_n = \sum_{i=1}^n x_i^2 y_i;$$

$$d) n \cdot 2 = 2 + 2 + \dots + 2 = \sum_{i=1}^n 2;$$

$$e) \sqrt{(x_1 - 2)^2 + (x_2 - 2)^2 + (x_3 - 2)^2 + \dots + (x_n - 2)^2} = \sqrt{\sum_{i=1}^n (x_i - 2)^2};$$

$$f) \frac{y_1^2 + y_2^2 + y_3^2}{3} - \left(\frac{y_1 + y_2 + y_3}{3} \right)^2 = \frac{\sum_{i=1}^3 y_i^2}{3} - \frac{1}{3} (\sum_{i=1}^3 y_i)^2;$$

Príklad 4

Nech je daná nasledujúca tabuľka hodnôt

x_i	n_i	y_i
1	2	3
3	3	4
4	5	5
5	7	3
7	1	6

Vypočítajte nasledujúce súčty: a) $\sum_{i=1}^n x_i$, b) $\sum_{i=1}^n x_i^2$, c) $\sum_{i=1}^n x_i n_i$, d) $\sum_{i=1}^n (x_i - y_i)^2$, e) $\sum_{i=1}^n (x_i - \bar{x})^2$, kde $\bar{x}$ je aritmetický priemer hodnôt x_i .

Riešenie.

Pomocou výpočtovej tabuľky dostaneme:

	x_i	n_i	y_i	x_i^2	$x_i n_i$	$(x_i - y_i)^2$	$(x_i - \bar{x})^2$
	1	2	3	1	2	4	9
	3	3	4	9	8	1	1
	4	5	5	16	20	1	0
	5	7	3	25	35	4	1
	7	1	6	49	7	1	9
Σ :	20			100	73	11	20

$\bar{x} = \frac{20}{5} = 4$

Zostrojenie takejto tabuľky je v Exceli veľmi efektívne, nakoľko tabuľkový kalkulačtor nám umožňuje vytvárať vlastné vzorce a ich jednoduché kopírovanie ťahaním myšou. Vzorce sú predpisy, ktoré vykonávajú výpočty so zadanými hodnotami v bunkách pracovných hárkov. Každý vzorec musí začínať znakom rovnosti = . Pri jeho zadávaní väčšinou využívame adresáciu buniek.

Ako sme realizovali výpočet v Exceli je naznačené na nasledujúcom obr.

v prvom riadku vytvoríme vzorce na výpočet požadovaných hodnôt

celý riadok označíme a nami vytvorené vzorce kopírujeme do ďalších riadkov tabuľky

vzorec na výpočet aritmetického priemeru hodnôt x_i

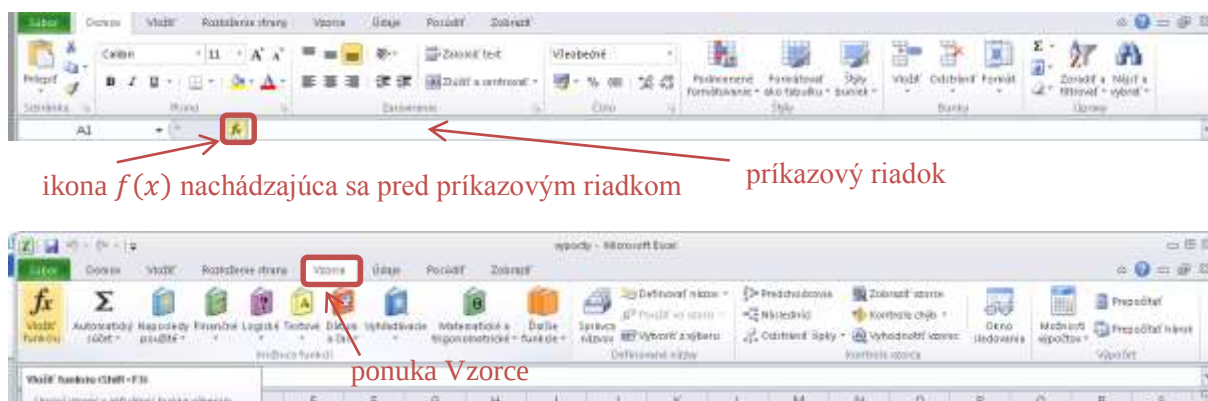
v tabuľke s kopírovanými vzorcami sa adresa kopírovaných buniek mení podľa smeru kopírovania (relatívna adresácia)

odkaz na bunku s aritmetickým priemerom sa v kopírovaných vzorcoch nemení (absolútna adresácia)

Obr. 1 Výpočet v Exceli – zadávanie vzorcov a ich kopírovanie

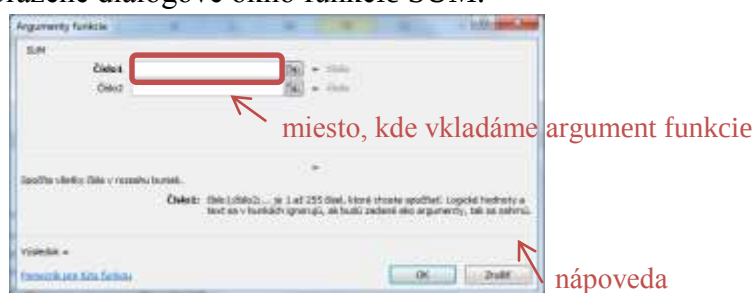
Na výpočet súčtov môžeme použiť preddefinované vzorce – v Exceli sú označené ako funkcie. Ich zadávanie je možné viacerými spôsobmi:

Prvý spôsob predstavuje vloženie funkcie prostredníctvom ponúkaného zoznamu preddefinovaných vzorcov, kedy nás program v dialógovom okne (obr. 3) navádza, aké argumenty je nevyhnutné vo funkcii zadať, aby bol výpočet správny. Zoznam ponúkaných funkcií sa zobrazí po kliknutí na ikonu , ktorá sa nachádza pred príkazovým riadkom alebo voľbou ponuky Vzorce v hlavnom menu súboru, obr. 2.



Obr. 2 Aktivácia zoznamu ponúkaných funkcií v Exceli

Na nasledujúcom obr. 3 je zobrazené dialógové okno funkcie SUM.



Obr. 3 Dialógové okno funkcie SUM

Druhým spôsobom je priame zadávanie funkcie do bunky alebo do príkazového riadku, pričom funkcia, rovnako ako všetky vzorce, musí začínať znakom = . Týmto spôsobom môžeme zadať nielen ľubovoľnú funkciu, ktorá je v Exceli preddefinovaná, ale preddefinované funkcie môžeme tiež použiť ako súčasť vlastného predpisu, vytvoriť vnorené funkcie a iné, obr. 4 a 5.

Pri zadávaní argumentov funkcií oboma spôsobmi je efektívne odvolávať sa bunky, ktoré obsahujú vstupné údaje pomocou ich adries. Adresa každej bunky je totiž určená pozíciou bunky v hárku pomocou súradnice riadku a stĺpca.



Obr. 4 Zadávanie funkcie SUM v Exceli bez použitia dialógového okna

		D3		f _x	=SQRT(2)*ABS(A1)	
		A	B	C	D	E
1		1				
2		3				
3		4			1,414214	

Obr. 5 K vytváraníu vlastných vzorcov

Súčtový riadok získame súčtom nami vytvorených pomocných výpočtov, obr. 6.

	A	B	C	D	E	F	G	H
1		x_i	n_i	y_i	x_i^2	$x_i n_i$	$(x_i - y_i)^2$	$(x_i - \bar{x})^2$
2		1	2	3	1	2	4	9
3		3	3	4	9	9	1	1
4		4	5	5	16	20	1	0
5		5	7	3	25	35	4	1
6		7	1	6	49	7	1	9
7		Σ: =SUM(B2:B6) =SUM(C2:C6) =SUM(D2:D6) =SUM(E2:E6) =SUM(F2:F6) =SUM(G2:G6) =SUM(H2:H6)						
8		$\bar{x} = 4$						
7		Σ:	20	18	21	100	73	11

Obr. 6 Výpočet súčtového riadku pomocou súčtov pomocných výpočtov

Hodnoty súčtov požadovaných v zadání príkladu však vieme v Exceli vypočítať aj priamo bez pomocných výpočtov pomocou vybraných funkcií. Ich syntax a popis sú nasledovné:

=SUM(číslo1;číslo2;...) vypočíta súčet zadaných čísiel;

=SUMSQ(číslo1;číslo2;...) vráti hodnotu súčtu druhých mocnín argumentov;

=SUMPRODUCT(pole1;pole2;...) vráti hodnotu súčtu súčinov zodpovedajúcich prvkov poľa;

=SUMXMY2(pole_x;pole_y) vráti súčet rozdielov druhých mocnín hodnôt dvoch zodpovedajúcich polí;

=DEVSQ(číslo1;číslo2;....) vráti súčet druhých mocnín odchýlok údajových bodov od strednej hodnoty vzorky;

=AVERAGE(číslo1;číslo2;....) vráti priemernú hodnotu argumentov.

	A	B	C	D	E	F	G	H
1		x_i	n_i	y_i	x_i^2	$x_i n_i$	$(x_i - y_i)^2$	$(x_i - \bar{x})^2$
2		1	2	3				
3		3	3	4				
4		4	5	5				
5		5	7	3				
6		7	1	6				
7		Σ: =SUM(B2:B6) =SUMSQ(B2:B6) =SUMPRODUCT(B2:B6;C2:C6) =SUMXMY2(B2:B6;D2:D6) =DEVSQ(B2:B6)						
7		Σ:	20		100	73	11	20

Obr. 7 Výpočet súčtového riadku pomocou vybraných funkcií

a) $\sum_{i=1}^n x_i = \text{SUM}(1;3;4;5;7)=20$, b) $\sum_{i=1}^n x_i^2 = \text{SUMSQ}(1;3;4;5;7) = 100$, c) $\sum_{i=1}^n x_i n_i = \text{SUMPRODUCT}(\text{pole } x_i; \text{pole } n_i) = 73$, d) $\sum_{i=1}^n (x_i - y_i)^2 = \text{SUMXMY2}(\text{pole } x_i; \text{pole } y_i) = 11$, e) $\sum_{i=1}^n (x_i - \bar{x})^2 = \text{DEVSQ}(1;3;4;5;7) = 20$, kde $\bar{x} = \text{AVERAGE}(1;3;4;5;7)$.

Matice a determinanty

Definícia. Nech m, n sú prirodzené čísla. Systém $m \cdot n$ prvkov množiny $M \subset R$, usporiadaných do m riadkov a n stĺpcov, sa nazýva matica typu $m \times n$. Označuje sa $A = (a_{ij}), i = 1, 2, \dots, m, j = 1, 2, \dots, n$ alebo

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & \dots & a_{mn} \end{pmatrix}.$$

Prvok a_{ij} je v i -tom riadku a j -tom stĺpci.

Matica typu $n \times n$ sa nazýva štvorcovou maticou stupňa n .

Definícia. Nech matica $A = (a_{ij})$ je typu $m \times n$. Matica $A^T = (a_{ij}^T)$ typu $n \times m$ s prvkami

$a_{ij}^T = a_{ji}, i = 1, 2, \dots, m, j = 1, 2, \dots, n$ sa nazýva transponovanou maticou k matici A .

Maticu A^T dostaneme z matice A zámenou riadkov za stĺpce, zachovajúc pritom ich poradie.

Definícia. Štvorcová matica A stupňa n , v ktorej $a_{ii} = 1$ a $a_{ij} = 0$ pre každé $i \neq j; i, j = 1, 2, \dots, n$ sa nazýva jednotková matica stupňa n a označuje sa E .

Definícia. (Súčet matíc) Súčtom matíc $A = (a_{ij}), B = (b_{ij})$, typu $m \times n$ nazývame maticu $C = (c_{ij})$ typu $m \times n$ s prvkami $c_{ij} = a_{ij} + b_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, n$. Súčet matíc A a B zapisujeme $C = A + B$.

Definícia. (Násobenie matice číslom) Súčinom reálneho čísla k a matice $A = (a_{ij})$ typu $m \times n$ nazývame maticu $C = (c_{ij})$ typu $m \times n$ s prvkami $c_{ij} = ka_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, n$. Zapisujeme $C = kA$.

Definícia. (Násobenie matíc) Nech matica $A = (a_{ij})$ je typu $m \times n$ a $B = (b_{ij})$, typu $n \times p$.

Súčinom matíc A a B (v tomto poradí) je matica $C = (c_{ij})$ typu $m \times p$ s prvkami

$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}, i = 1, 2, \dots, m, j = 1, 2, \dots, p$. Píšeme $C = AB$.

Pre násobenie matíc neplatí vo všeobecnosti komutatívny zákon.

Definícia. Determinantom štvorcovej matice A stupňa n sa nazýva číslo, označené $|A|$, ktoré sa rovná

$|A| = a_{11}$, ak matica A je stupňa 1, t.j. $A = (a_{11})$

$|A| = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| + \dots + (-1)^{n+1}a_{1n}|A_{1n}|$, ak matica A je stupňa $n > 1$, kde $|A_{ij}|$ je determinant matice, ktorá vznikne z A vynechaním i -teho riadku a j -teho stĺpca.

Podľa tejto rekurentnej definície determinant matice druhého stupňa je

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}|A_{11}| - a_{12}|A_{12}| = a_{11}a_{22} - a_{12}a_{21}$$

a determinant matice tretieho stupňa je číslo

$$\begin{aligned} |A| &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - \\ &a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = \\ &a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

Posledný vzťah môžeme dostať podľa tzv. Sarrusovho pravidla: K stĺpcom matice A pripíšeme prvé dva stĺpce a utvoríme po troch prvkoch z každého riadku a stĺpca v smere hlavnej diagonály so znamienkom (+) a v smere vedľajšej diagonály so znamienkom (–) a výsledok sčítame.

Definícia. Nech A je štvorcová matica, ktorej determinant sa nerovná nule. Maticu A^{-1} , pre ktorú platí $AA^{-1} = A^{-1}A = E$ nazývame inverznou maticou k matici A .

Príklad 5

Nájdite maticu $C = 0,2A + 1,3B - 4A^T$, kde $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 3 & 2 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 3 & 5 \\ 0 & 5 & 2 \end{pmatrix}$.

Riešenie.

Podľa definície o súčine matice a konštanty máme $0,2A = \begin{pmatrix} 0,2 & 0,2 & 0,2 \\ 0,4 & 0 & 0,2 \\ 0,6 & 0,4 & 0,8 \end{pmatrix}$, $1,3B = \begin{pmatrix} 1,3 & 0 & 1,3 \\ 2,6 & 3,9 & 6,5 \\ 0 & 6,5 & 2,6 \end{pmatrix}$. Trasponovaná matica k matici A bude $A^T = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 2 \\ 1 & 1 & 4 \end{pmatrix}$, následne

$$-4A^T = \begin{pmatrix} -4 & -8 & -12 \\ -4 & 0 & -8 \\ -4 & -4 & -16 \end{pmatrix}.$$

Hľadanú maticu nájdeme ako súčet matic

$$C = 0,2A + 1,3B - 4A^T = \begin{pmatrix} -2,5 & -7,8 & -10,5 \\ -1 & 3,9 & -1,3 \\ -3,4 & 2,9 & -12,6 \end{pmatrix}.$$

Celý výpočet v Exceli môžeme vidieť na nasledujúcom obr. 8.

	A	B	C	D	E	F	G	H	I
1									
2	A=	1	1	1		B=	1	0	1
3		2	0	1			2	3	5
4		3	2	4			0	5	2
5									
6	0,2A=	=0,2*B2	=0,2*C2	=0,2*D2		1,3B=	=1,3*G2	=1,3*H2	=1,3*I2
7		=0,2*B3	=0,2*C3	=0,2*D3			=1,3*G3	=1,3*H3	=1,3*I3
8		=0,2*B4	=0,2*C4	=0,2*D4			=1,3*G4	=1,3*H4	=1,3*I4
9									
10		=TRANSPOSE(B2:D4)							
11		TRANSPOSE(mat)							
12									
13									
14	-4A ^T =	=-4*B10	=-4*C10	=-4*D10					
15		=-4*B11	=-4*C11	=-4*D11					
16		=-4*B12	=-4*C12	=-4*D12					
17									
18	C=	=B6+G6+B14	=C6+H6+C14	=D6+I6+D14					
19		=B7+G7+B15	=C7+H7+C15	=D7+I7+D15					
20		=B8+G8+B16	=C8+H8+C16	=D8+I8+D16					

Obr. 8 K výpočtu v Exceli

Pri výpočte súčinu matice a konštanty a súčte matic určujeme hodnoty jednotlivých prvkov novej matice postupne po jednotlivých bunkách. Využívame pritom kopírovanie vytvorených vzorcov do požadovaných buniek ťahaním myšou. Podľa smeru kopírovania sa automaticky menia odkazy na bunky.

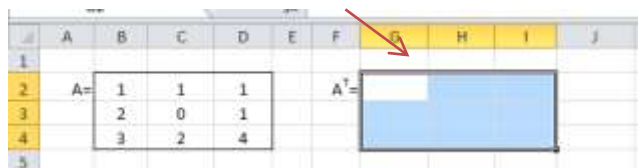
	A	B	C	D	E
1					
2	A=	1	1	1	
3		2	0	1	
4		3	2	4	
5					
6	0,2A=	=0,2*B2			
7					
8					

V prvej bunke matice definujeme vzorec na výpočet hodnoty daného prvku, ostatné bunky matice následne vyplníme kopírovaním.

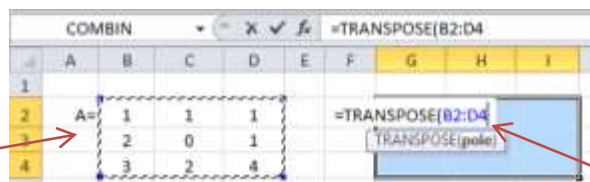
Pri určovaní transponovanej matice príkaz TRASPOSE vyžaduje, aby sme s maticou pracovali ako s poľom, t.j. so súvislou plochou buniek. Pri takomto type funkcií, kedy výstupom nie je iba jedna bunka, ale súvislé pole buniek, je potrebné príkaz vždy ukončiť trojkombináciou kláves CTRL+Shift+Enter.

Pri zadávaní funkcie je dôležité dodržať nasledovnú postupnosť krokov:

1. Vyznačíme oblasť, kde má program výslednú maticu napísať.



3. Ako argument funkcie označíme celú maticu A, ku ktorej hľadáme transponovanú maticu.



4. Výpočet potvrdíme pomocou CTRL+Shift+Enter !

Obr. 9 Výpočet transponovanej funkcie v Exceli

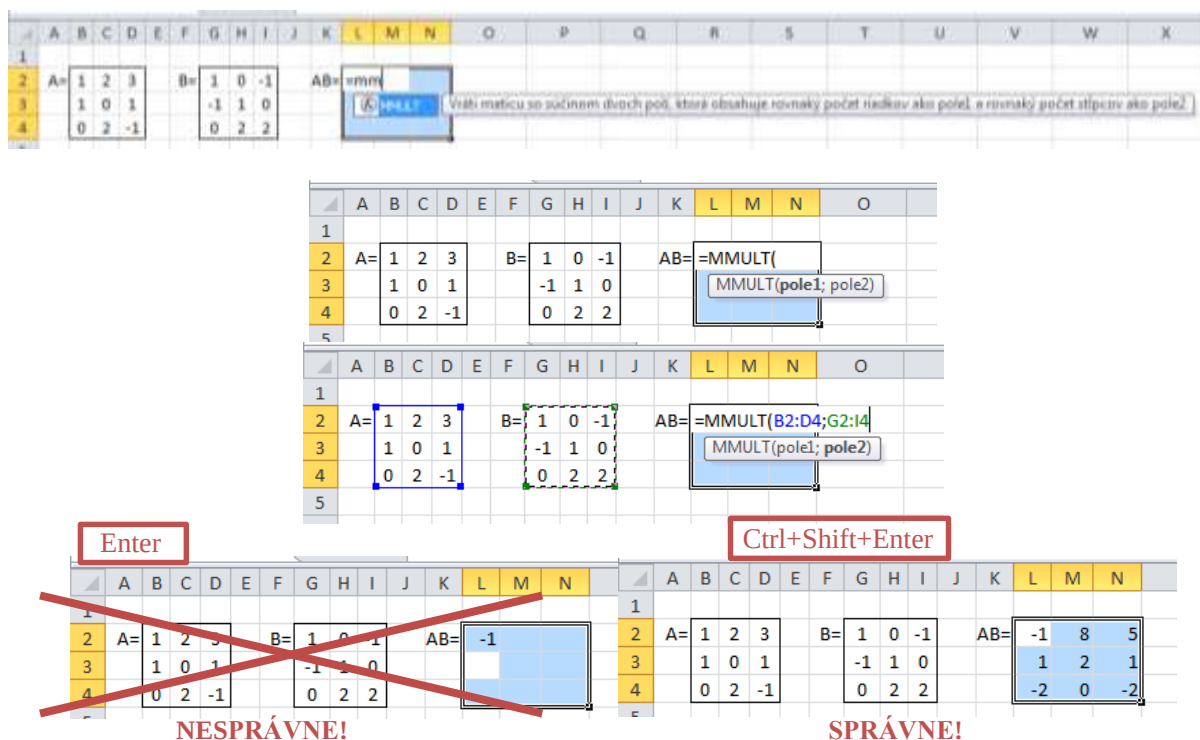
Príklad 6 Dané sú matice $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 2 & -1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & 2 & 2 \end{pmatrix}$. Nájdite ich súčin.

Riešenie.

Nájdeme súčin AB . Podľa definície súčinu matíc dostaneme $AB =$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & 2 & 2 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot (-1) + 3 \cdot 0 & 1 \cdot 0 + 2 \cdot 1 + 3 \cdot 2 & 1 \cdot (-1) + 2 \cdot 0 + 3 \cdot 2 \\ 1 \cdot 1 + 0 \cdot (-1) + 1 \cdot 0 & 1 \cdot 0 + 0 \cdot 1 + 1 \cdot 2 & 1 \cdot (-1) + 0 \cdot 0 + 1 \cdot 2 \\ 0 \cdot 1 + 2 \cdot (-1) + (-1) \cdot 0 & 0 \cdot 0 + 2 \cdot 1 + (-1) \cdot 2 & 0 \cdot (-1) + 2 \cdot 0 + (-1) \cdot 2 \end{pmatrix} = \begin{pmatrix} -1 & 8 & 5 \\ 1 & 2 & 1 \\ -2 & 0 & -2 \end{pmatrix}.$$

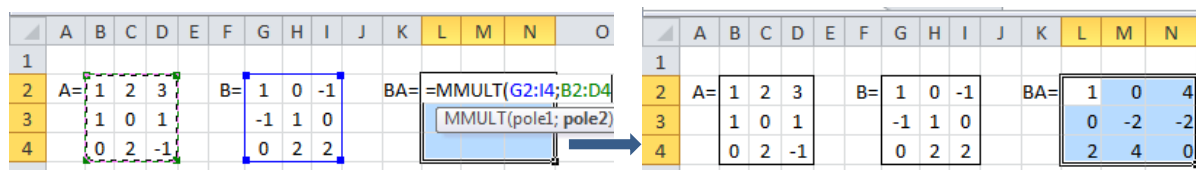
V Exceli najprv vyznačíme oblasť, kam umiestnime inverznú maticu typu 3×3 (3 riadky a 3 stĺpce). Použijeme funkciu =MMULT(pole A; pole B) a potvrdíme trojkombináciou kláves Ctrl + Shift + Enter. Ak by sme potvrdili iba klávesou Enter alebo OK, tak funkcia vypíše iba prvý prvok hľadanej matice, obr.10.



Obr. 10 Násobenie matic v Exceli

$$\text{Vypočítaný súčin } BA = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 4 \\ 0 & -2 & -2 \\ 2 & 4 & 0 \end{pmatrix}.$$

Výpočet v Exceli môžeme vidieť na nasledujúcom obr.11.

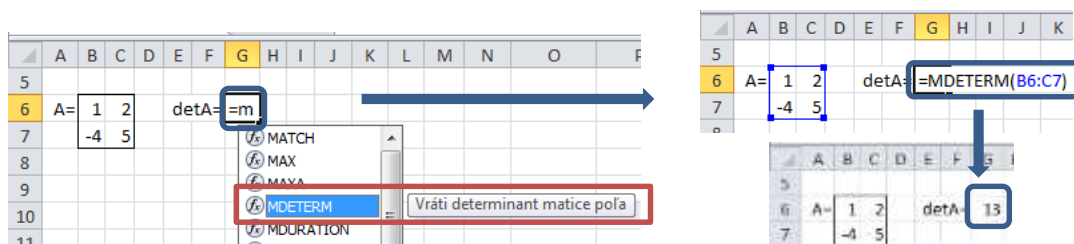


Obr. 11 Násobenie matic v Exceli – funkciu MMULT potvrdíme pomocou Ctrl+Shift+Enter

Príklad 7 Daná je matica $A = \begin{pmatrix} 1 & 2 \\ -4 & 5 \end{pmatrix}$. Vypočítajte jej determinant.

Riešenie.

Podľa definície determinantu matice druhého stupňa, determinant sa rovná $|A| = \begin{vmatrix} 1 & 2 \\ -4 & 5 \end{vmatrix} = 1 \cdot 5 - (-4) \cdot 2 = 13$. Výpočet pomocou Excelu dostaneme $|A| = \text{MDETERM}(\text{pole } A) = 13$, obr. 12.



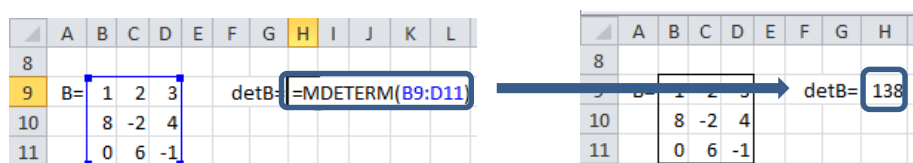
Obr. 12 Výpočet determinantu matice v Exceli

Príklad 8 Daná je matrica $B = \begin{pmatrix} 1 & 2 & 3 \\ 8 & -2 & 4 \\ 0 & 6 & -1 \end{pmatrix}$. Vypočítajte jej determinant.

Riešenie. Pre výpočet determinantu matice tretieho stupňa použijeme Sarrusovo pravidlo

$$|B| = \begin{vmatrix} 1 & 2 & 3 \\ 8 & -2 & 4 \\ 0 & 6 & -1 \end{vmatrix} = 1 \cdot (-2) \cdot (-1) + 2 \cdot 4 \cdot 0 + 3 \cdot 8 \cdot 6 - 0 \cdot (-2) \cdot 3 - 6 \cdot 4 \cdot 1 - (-1) \cdot 8 \cdot 2 = 138.$$

V Exceli (obr. 13) determinant dostaneme pomocou funkcie MDETERM(pole B) = 138.

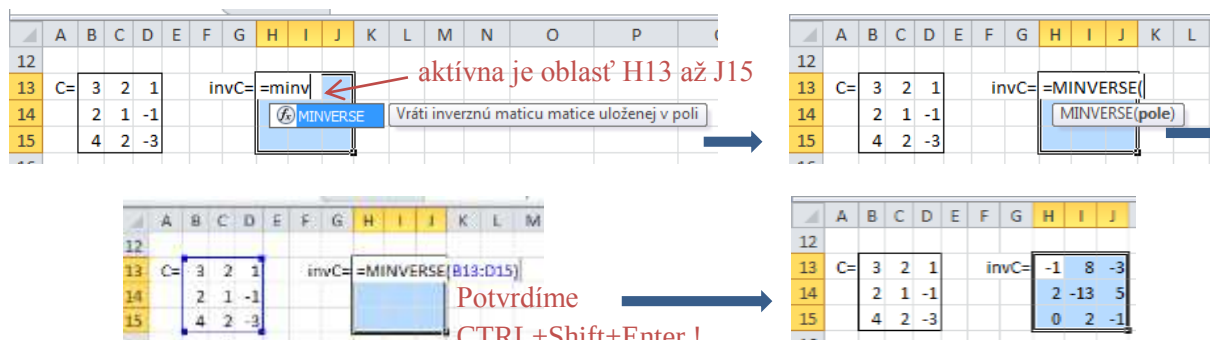


Obr. 13 Výpočet determinantu matice

Príklad 9 Daná je matrica $C = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 1 & -1 \\ 4 & 2 & -3 \end{pmatrix}$. Nájdite k nej inverznú maticu.

Riešenie.

Inverznú maticu C^{-1} nájdeme v Exceli. Vyznačíme oblasť, kam umiestnime inverznú maticu typu 3×3 (3 riadky a 3 stĺpce). Použijeme funkciu =MINVERSE(pole C), potvrdíme trojkombináciou kláves Ctrl+Shift+Enter.



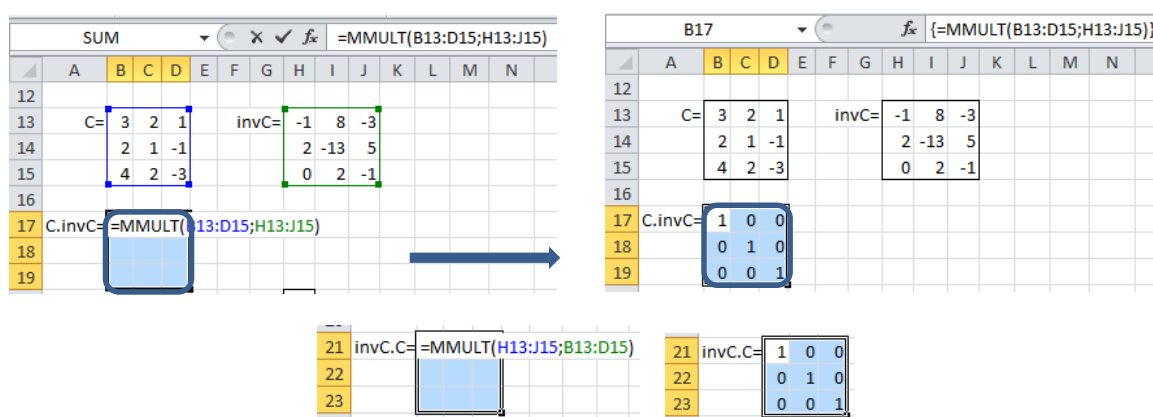
Obr. 14 Výpočet inverznej matice

$$\text{Inverzná matica } C^{-1} = \begin{pmatrix} -1 & 8 & -3 \\ 2 & -13 & 5 \\ 0 & 2 & -1 \end{pmatrix}.$$

Presvedčme sa, že podľa uvedenej definície inverznej matice, matica C^{-1} je skutočne inverzná k danej matici C . K tomu potrebujeme nájsť súčin týchto dvoch matíc v ľubovoľnom poradí, obr. 15:

$$CC^{-1} = \text{MMULT}(\text{pole } C; \text{pole } C^{-1}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E$$

$$C^{-1}C = \text{MMULT}(\text{pole } C^{-1}; \text{pole } C) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E$$



Obr. 15 Súčin matic $CC^{-1} = C^{-1}C = E$

System lineárnych algebraických rovníc

Definícia. Systémom m lineárnych algebraických rovníc s n neznámymi rozumieme nasledujúci systém rovníc:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n &= b_m \end{aligned}$$

kde x_i sú neznáme, a_{ij} sú koeficienty systému a b_i sú absolútne členy alebo pravé strany systému. K danému systému možno priradiť nasledujúce matice:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \text{ typu } m \times n, \text{ ktorú nazývame maticou systému a matice}$$

$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ a $B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$ typu $n \times 1$ resp. $m \times 1$. Maticu X nazývame vektor neznámych

a B vektor absolútnych členov. Potom systém môžeme zapísať v maticovom tvare

$$AX = B.$$

Definícia. Riešením systému lineárnych algebraických rovníc nazývame každý vektor

$C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$, pre ktorý platí $AC = B$. Riešiť systém znamená nájsť všetky jeho riešenia.

Kvôli jednoduchosti vektor riešení budeme písať aj v tvare riadkového vektora $C = (c_1, c_2, \dots, c_n)$.

Veta. (Riešenie pomocou inverznej matice) Nech je daný systém n lineárnych algebraických rovníc s n neznámymi tvaru $AX = B$ (t.j. matica systému je štvorcová). Nech matica A je regulárna (t.j. determinant matice A je rôzny od nuly). Potom vektor $X = A^{-1}B$ je jediným riešením tohto systému.

Veta. (Crammerovo pravidlo)) Nech je daný systém n lineárnych algebraických rovníc s n neznámymi tvaru $AX = B$ (t.j. matica systému je štvorcová). Nech matica A je regulárna (t.j. $|A| \neq 0$). Potom daný systém má jediné riešenie tvaru $(x_1, x_2, \dots, x_n) = \left(\frac{D_1}{D}, \frac{D_2}{D}, \dots, \frac{D_n}{D}\right)$, kde $D = |A|$ je determinant matice systému a D_i je determinant matice, ktorá vznikne z matice A tak, že jej i -tý stĺpec nahradíme stĺpcom absolútnych členov.

Príklad 10 Pomocou inverznej matice riešte systém:

$$\begin{aligned} 3x_1 - 2x_2 + x_3 &= 11; \\ x_1 + x_2 - 3x_3 &= 7; \\ 1x_1 - 4x_2 - 3x_3 &= 10; \end{aligned}$$

Riešenie.

V Exceli (obr. 16) nájdeme determinant matice systému

$$|A| = \begin{vmatrix} 3 & -2 & 1 \\ 1 & 1 & -3 \\ 1 & -4 & -3 \end{vmatrix} = \text{MDETERM}(\text{pole matice}) = -50.$$

	A	B	C	D	E	F	G	H	I	J	K	L
28												
29	A=	3	-2	1	detA=	=MDETERM(B29:D31)						
30		1	1	-3		MDETERM(pole)						
31		1	-4	-3								

Obr. 16 Výpočet determinantu matice

Vypočítame inverznú maticu $A^{-1} = \text{MINVERSE}(\text{pole matice}) = \begin{pmatrix} 0,3 & 0,2 & -0,1 \\ 0 & 0,2 & -0,2 \\ 0,1 & -0,2 & -0,1 \end{pmatrix}$,

obr. 17.

	A	B	C	D	E	F	G	H	I	J	K	L
28												
29	A=	3	-2	1	invA=	=MINVERSE(B29:D31)						
30		1	1	-3								
31		1	-4	-3								

Obr. 17 Výpočet inverznej matice

Stĺpec neznámych nájdeme podľa vzťahu

$$X = A^{-1}B = \begin{pmatrix} 0,3 & 0,2 & -0,1 \\ 0 & 0,2 & -0,2 \\ 0,1 & -0,2 & -0,1 \end{pmatrix} \begin{pmatrix} 11 \\ 7 \\ 10 \end{pmatrix} = \text{MMULT}(\text{pole } A^{-1}; \text{pole } B) = \begin{pmatrix} 3,7 \\ -0,6 \\ -1,3 \end{pmatrix}.$$

	E	F	G	H	I	J	K	L	M	N	O	P	Q
28													
29	invA=	0,3	0,2	-0,1	B=	11	invA.B=	=MMULT(G29:I31;L29:L31)					
30		0	0,2	-0,2		7							
31		0,1	-0	-0,1		10							

Obr. 18 Výpočet súčinu matic

Príklad 11 Pomocou Crammerovej vety riešte systém:

$$2x_1 - 3x_2 + x_3 = 0$$

$$x_1 + 2x_2 - x_3 = 3$$

$$2x_1 + x_2 + x_3 = 12$$

Riešenie.

V Exceli nájdeme determinant systému

$$D = \begin{vmatrix} 2 & -3 & 1 \\ 1 & 2 & -1 \\ 2 & 1 & 1 \end{vmatrix} = \text{MDETERM}(\text{pole matice}) = 12.$$

Vypočítame determinanty:

$$D_1 = \begin{vmatrix} 0 & -3 & 1 \\ 3 & 2 & -1 \\ 12 & 1 & 1 \end{vmatrix} = \text{MDETERM}(\text{pole matice}) = 24;$$

$$D_2 = \begin{vmatrix} 2 & 0 & 1 \\ 1 & 3 & -1 \\ 2 & 12 & 1 \end{vmatrix} = \text{MDETERM}(\text{pole matice}) = 36;$$

$$D_3 = \begin{vmatrix} 2 & -3 & 0 \\ 1 & 2 & 3 \\ 2 & 1 & 12 \end{vmatrix} = \text{MDETERM}(\text{pole matice}) = 60.$$

	A	B	C	D	E	F	G	H	I	J	K
32											
33	D=	2	-3	1		detD=	=MDETERM(B33:D35)				
34		1	2	-1							
35		2	1	1							
36											
37	D1=	0	-3	1		detD1=	=MDETERM(B37:D39)				
38		3	2	-1							
39		12	1	1							
40											
41	D1=	2	0	1		detD2=	=MDETERM(B41:D43)				
42		1	3	-1							
43		2	12	1							
44											
45	D1=	2	-3	0		detD3=	=MDETERM(B45:D47)				
46		1	2	3							
47		2	1	12							

	A	B	C	D	E	F	G	H
32								
33	D=	2	-3	1		detD=	12	
34		1	2	-1				
35		2	1	1				
36								
37	D1=	0	-3	1		detD1=	24	
38		3	2	-1				
39		12	1	1				
40								
41	D1=	2	0	1		detD2=	36	
42		1	3	-1				
43		2	12	1				
44								
45	D1=	2	-3	0		detD3=	60	
46		1	2	3				
47		2	1	12				

Obr. 19 Výpočet determinantov

Riadok neznámych dostaneme zo vzťahu $(x_1, x_2, x_3) = \left(\frac{D_1}{D}, \frac{D_2}{D}, \frac{D_3}{D}\right) = \left(\frac{24}{12}, \frac{36}{12}, \frac{60}{12}\right) = (2, 3, 5)$

x1=	=G37/\$G\$33	x2=	=G41/\$G\$33	x3=	=G45/\$G\$33
-----	--------------	-----	--------------	-----	--------------

x1=	2	x2=	3	x3=	5
-----	---	-----	---	-----	---

Obr. 20 Výpočet neznámych

Príklad 12

Nájdite maticu X , pre ktorú platí $AX = B$, ak $A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 3 & 5 \\ 3 & 5 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 7 & 1 & 1 \end{pmatrix}$.

Riešenie. Maticovú rovnicu $AX = B$ vynásobím zľava inverznou maticou A^{-1} a dostaneme

$$A^{-1}AX = A^{-1}B$$

$$EX = A^{-1}B$$

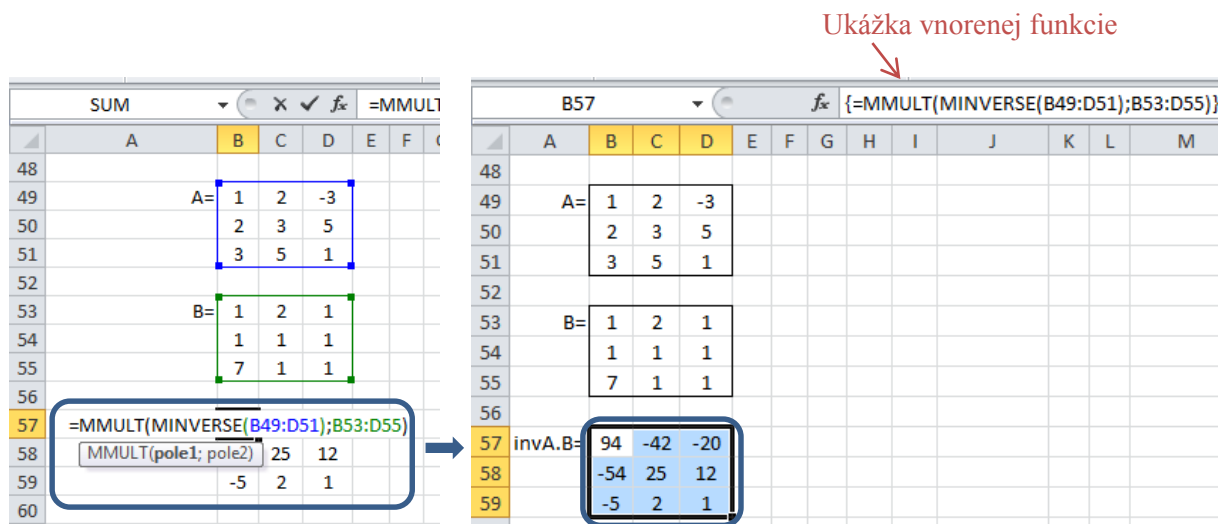
$$X = A^{-1}B$$

V Exceli vypočítame inverznú maticu k matici A (ktorá existuje, lebo A je regulárna, jej determinant je $|A| = \text{MDETERM}(\text{pole matice}) = 1$).

$$A^{-1} = \text{MINVERSE}(\text{pole matice}) = \begin{pmatrix} -22 & -17 & 19 \\ 13 & 10 & -11 \\ 1 & 1 & -1 \end{pmatrix}.$$

Potom neznáma matica $X = A^{-1}B = \begin{pmatrix} -22 & -17 & 19 \\ 13 & 10 & -11 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 7 & 1 & 1 \end{pmatrix} = \text{MMULT}(\text{pole } A^{-1}; \text{pole } B) = \begin{pmatrix} 94 & -42 & -20 \\ -54 & 25 & 12 \\ -5 & 2 & 1 \end{pmatrix}.$

Po pochopení princípu možno použiť aj skrátенý výpočet pomocou vnorenej funkcie, ktorý môžeme vidieť na nasledujúcom obr. 21.



Obr. 21 Skrátенý výpočet v Exceli

Príklad 13 Nájdite maticu X , pre ktorú platí $XA = B$, ak

$$A = \begin{pmatrix} 1 & 1 & -2 \\ 0 & 1 & 1 \\ -1 & 2 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -1 & 0 \\ -3 & 3 & 8 \\ 0 & 5 & 1 \end{pmatrix}.$$

Riešenie.

Maticovú rovnicu $XA = B$ vynásobím sprava inverznou maticou A^{-1} a dostaneme

$$\begin{aligned} XAA^{-1} &= BA^{-1} \\ XE &= BA^{-1} \\ X &= BA^{-1} \end{aligned}$$

V Exceli vypočítame inverznú maticu k matici A (ktorá existuje, lebo A je regulárna, jej determinant nie je nulový, $|A| = \text{MDETERM}(\text{pole matice}) = -4$).

$$A^{-1} = \text{MINVERSE}(\text{pole matice}) = \begin{pmatrix} 0,25 & 1,25 & -0,75 \\ 0,25 & 0,25 & 0,25 \\ -0,25 & 0,75 & -0,25 \end{pmatrix}.$$

Potom neznáma matica $X = BA^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ -3 & 3 & 8 \\ 0 & 5 & 1 \end{pmatrix} \begin{pmatrix} 0,25 & 1,25 & -0,75 \\ 0,25 & 0,25 & 0,25 \\ -0,25 & 0,75 & -0,25 \end{pmatrix} =$

$$\text{MMULT}(\text{pole } B; \text{pole } A^{-1}) = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 3 & 1 \\ 1 & 2 & 1 \end{pmatrix}.$$

Skrátený výpočet príkladu v Exceli je na nasledujúcom obr. 22.

The image shows two Excel spreadsheets. The left spreadsheet has columns A through F and rows 61 through 72. It defines matrix A in cells B62:D64 and matrix B in cells B66:D68. A formula in cell B70, highlighted with a blue box, is `=MMULT(B66:D68;MINVERSE(B62:D64))`. The right spreadsheet has columns A through M and rows 61 through 72. It defines matrix A in cells B62:D64 and matrix B in cells B66:D68. A formula in cell B70, also highlighted with a blue box, is `=MMULT(B66:D68;MINVERSE(B62:D64))`. Both spreadsheets show the result of the matrix multiplication in cells B71:D72.

Obr. 22 Skrátený výpočet v Exceli

Príklad 14 Nájdite maticu X , pre ktorú platí $AX = B$, ak

$$\begin{aligned} 2x_1 + x_2 - x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 1 \\ x_1 - 2x_2 + x_3 &= -2 \\ x_1 + x_2 + x_3 &= 4 \end{aligned}$$

Riešenie. Maticovú rovnicu $AX = B$ vynásobíme zľava transponovanou maticou A^T , aby sme dostali štvorcovú maticu $A^T A$.

$$A^T AX = A^T B$$

Túto rovnicu vynásobíme zľava inverznou maticou $(A^T A)^{-1}$, o ktorej predpokladáme, že existuje.

$$\begin{aligned} (A^T A)^{-1}(A^T A)X &= (A^T A)^{-1}A^T B \\ EX &= (A^T A)^{-1}A^T B \\ X &= (A^T A)^{-1}A^T B \end{aligned}$$

$$\text{Vypočítame } A^T = \begin{pmatrix} 2 & 1 & -1 \\ 1 & -1 & 2 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{pmatrix}^T = \text{TRANSPOSE}(\text{pole matice}) = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & -1 & -2 & 1 \\ -1 & 2 & 1 & 1 \end{pmatrix}.$$

Nájdeme súčin matíc

$$A^T A = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & -1 & -2 & 1 \\ -1 & 2 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & -1 \\ 1 & -1 & 2 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \text{MMULT}(\text{pole } A^T; \text{pole } A) = \begin{pmatrix} 7 & 0 & 2 \\ 0 & 7 & -4 \\ 2 & -4 & 7 \end{pmatrix}.$$

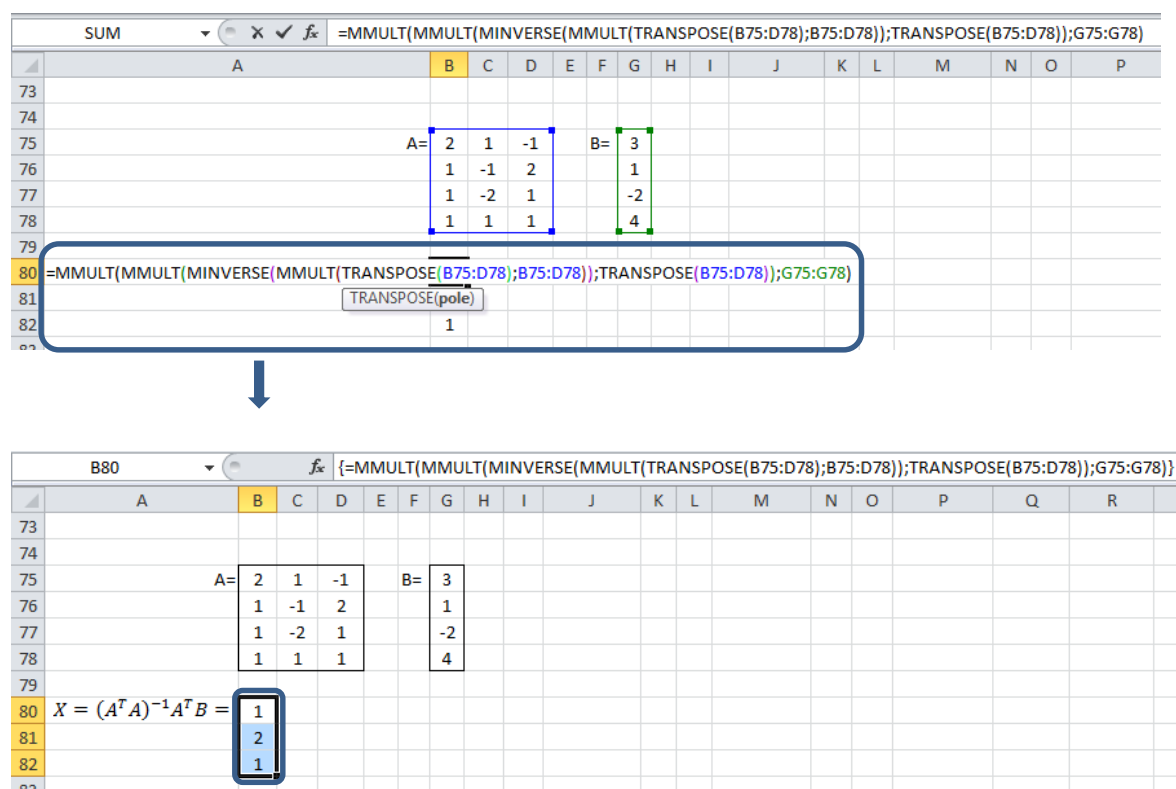
K tejto štvorcovej matici nájdeme inverznú

$$(A^T A)^{-1} = \text{MINVERSE}(\text{pole matice}) = \begin{pmatrix} 0,162562 & -0,03941 & -0,06897 \\ -0,03941 & 0,221675 & 0,137931 \\ -0,06897 & 0,137931 & 0,241379 \end{pmatrix}.$$

Neznámu maticu X dostaneme ako súčin troch matíc

$$X = (A^T A)^{-1} A^T B = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

Skrátený výpočet v Exceli je uvedený na nasledujúcom obr. 23.



Obr. 23 Skrátený výpočet v Exceli

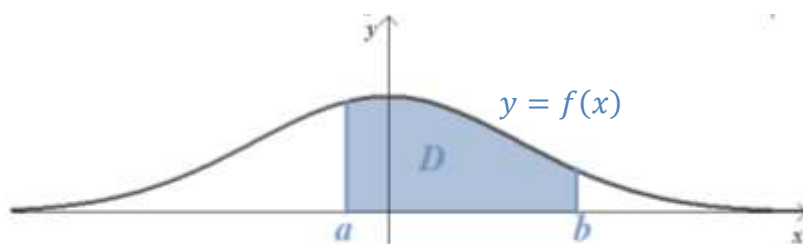
Určitý integrál

Veta. (Newton-Leibnizov vzorec) Nech je funkcia $f: y = f(x)$ integrovateľná na intervale $\langle a, b \rangle$. Nech je funkcia $F(x)$ spojitá na $\langle a, b \rangle$ a nech je primitívnou funkciou k funkcii f na $\langle a, b \rangle$. Potom platí

$$\int_a^b f(x) dx = [F(x)]_a^b = F(b) - F(a).$$

Veta. Nech je funkcia f integrovateľná na intervale $\langle a, b \rangle$ a $f(x) \geq 0$ pre každé $x \in \langle a, b \rangle$. Potom platí $\int_a^b f(x) dx \geq 0$.

Veta. Ak funkcia f je spojitá a nezáporná na intervale $\langle a, b \rangle$, potom určitý integrál $\int_a^b f(x) dx$ určuje plošný obsah krivočiareho lichobežníka $D = \{[x, y] \in E_2: a \leq x \leq b, 0 \leq y \leq f(x)\}$.



Veta. (Substitučná metóda) Ak sa integrovateľná funkcia na intervale $\langle a, b \rangle$ dá vyjadriť v tvare $f(\varphi(x))\varphi'(x)$, kde funkcie φ a φ' sú spojité na $\langle a, b \rangle$ a zároveň funkcia f je spojitá v každom bode $t = \varphi(x)$, kde $x \in \langle a, b \rangle$, tak platí

$$\int_a^b f(\varphi(x))\varphi'(x) dx = \left| \begin{matrix} t = \varphi(x) \\ dt = \varphi'(x) dx \end{matrix} \right| = \int_{\varphi(a)}^{\varphi(b)} f(t) dt.$$

Veta. (Metóda per partes) Nech majú funkcie u, v na intervale $\langle a, b \rangle$ spojité derivácie u', v' . Potom platí

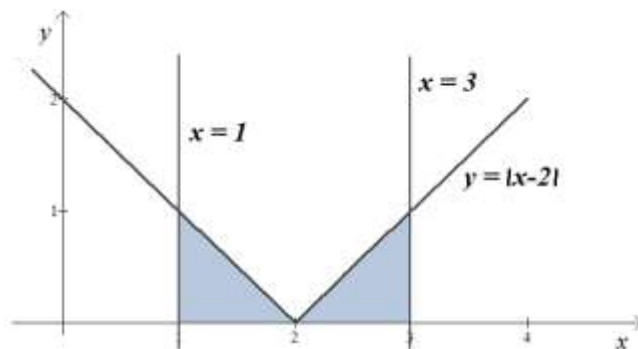
$$\int_a^b u(x)v'(x) dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x) dx.$$

Příklad 15

Pomocou určitého integrálu vypočítajte plošný obsah časti roviny, ohraničenej grafmi funkcií $f: y = 0$, $g: y = |x - 2|$, a priamkami $x = 1$ a $x = 3$.

Riešenie. Funkcia $f: y = 0$ je x -ová os. Funkciu g môžeme pre zjednodušenie našich výpočtov zapísať v tvare

$$g: y = \begin{cases} 2 - x, & \text{pre } x \leq 2; \\ x - 2, & \text{pre } x \geq 2; \end{cases}$$



Obr. 24 Časť roviny ohraničená zadanými funkciami

$$P_l = \int_1^2 (2-x) dx + \int_2^3 (x-2) dx = \left[2x - \frac{x^2}{2} \right]_1^2 + \left[\frac{x^2}{2} - 2x \right]_2^3 = \left(2 \cdot 2 - \frac{2^2}{2} \right) - \left(2 \cdot 1 - \frac{1^2}{2} \right) + \left(\frac{3^2}{2} - 2 \cdot 3 \right) - \left(\frac{2^2}{2} - 2 \cdot 2 \right) = 2 - \frac{3}{2} - \frac{3}{2} - (-2) = 1$$

Nevlastný integrál

Definícia. Nech je funkcia f definovaná na intervale $\langle a, \infty \rangle$ a integrovateľná na intervale $\langle a, b \rangle$ pre každé $b > a$. Ak existuje vlastná limita

$$\lim_{b \rightarrow \infty} \int_a^b f(x) dx,$$

nazveme ju nevlastný integrál funkcie f na intervale $\langle a, \infty \rangle$. Píšeme

$$\lim_{b \rightarrow \infty} \int_a^b f(x) dx = \int_a^{\infty} f(x) dx.$$

Definícia. Nech je funkcia f definovaná na intervale $(-\infty, b)$ a integrovateľná na intervale $\langle a, b \rangle$ pre každé $a < b$. Ak existuje vlastná limita

$$\lim_{a \rightarrow -\infty} \int_a^b f(x) dx,$$

nazveme ju nevlastný integrál funkcie f na intervale $(-\infty, b)$ a zapíšeme

$$\lim_{a \rightarrow -\infty} \int_a^b f(x) dx = \int_{-\infty}^b f(x) dx.$$

Ak v uvedených definíciách limita existuje, hovoríme, že nevlastný integrál konverguje. Ak limita neexistuje alebo je nevlastná, hovoríme, že nevlastný integrál diverguje.

Definícia. Nech je funkcia f definovaná na intervale $(-\infty, \infty)$. Nech c je ľubovoľné číslo. Ak existujú nevlastné integrály $\int_{-\infty}^c f(x) dx$ a $\int_c^{\infty} f(x) dx$, ich súčet nazveme nevlastný integrál funkcie f na intervale $(-\infty, \infty)$ a zapíšeme

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^c f(x) dx + \int_c^{\infty} f(x) dx.$$

Ak jeden z posledných integrálov diverguje, hovoríme, že nevlastný integrál funkcie f na intervale $(-\infty, \infty)$ diverguje. V opačnom prípade hovoríme, že konverguje.

Príklad 16 Vypočítajte plošný obsah rovinného útvaru ohraničeného krivkami $y = e^{-x}$, $y = 0$, $x \in (0, \infty)$.

Riešenie.
$$P = \int_0^{\infty} e^{-x} dx = \lim_{a \rightarrow \infty} \int_0^a e^{-x} dx = \lim_{a \rightarrow \infty} [-e^{-x}]_0^a = \lim_{a \rightarrow \infty} [-e^{-a} + e^0] = \lim_{a \rightarrow \infty} \left(1 - \frac{1}{e^a}\right) = 1 - 0 = 1.$$

Príklad 17 Vypočítajte integrál funkcie f na intervale $(-\infty, \infty)$, ak funkcia f je daná

$$\text{predpisom } f(x) = \begin{cases} 0 & x < -2 \\ x^2 + 1 & x \in \langle -2, 0 \rangle \\ 3 & x \in (0, 4) \\ 0 & x > 4 \end{cases}.$$

Riešenie.
$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{-2} f(x) dx + \int_{-2}^0 f(x) dx + \int_0^4 f(x) dx + \int_4^{\infty} f(x) dx = \int_{-\infty}^{-2} 0 dx + \int_{-2}^0 (x^2 + 1) dx + \int_0^4 3 dx + \int_4^{\infty} 0 dx = 0 + \left[\frac{x^3}{3} + x\right]_{-2}^0 + [3x]_0^4 + 0 = \frac{8}{3} + 2 + 12 = \frac{50}{3}$$

Príklad 18 Vypočítajte $\int_0^{\infty} \frac{1}{1+x^2} dx$.

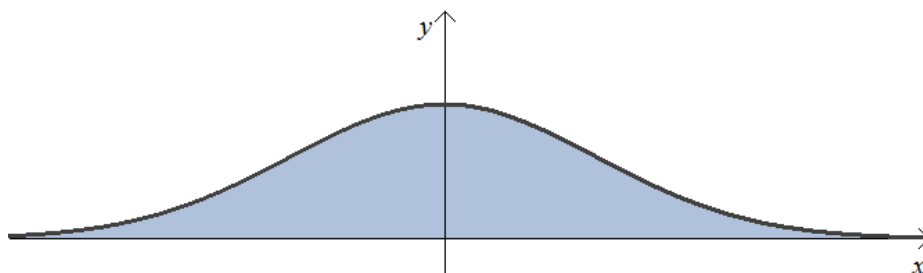
Riešenie.
$$\int_0^{\infty} \frac{1}{1+x^2} dx = \lim_{a \rightarrow \infty} \int_0^a \frac{1}{1+x^2} dx = \lim_{a \rightarrow \infty} [\arctg x]_0^a = \lim_{a \rightarrow \infty} (\arctg a - \arctg 0) = \frac{\pi}{2}.$$

Príklad 19 Vypočítajte $\int_0^{\infty} \frac{x}{1+x^2} dx$.

Riešenie.
$$\int_0^{\infty} \frac{x}{1+x^2} dx = \lim_{a \rightarrow \infty} \int_0^a \frac{x}{1+x^2} dx = \lim_{a \rightarrow \infty} \left[\frac{1}{2} \ln(1+x^2) \right]_0^a = \lim_{a \rightarrow \infty} \left(\frac{1}{2} \ln(1+a^2) - \frac{1}{2} \ln(1+0^2) \right) = \infty$$

Daný nevlastný integrál je divergentný.

Príklad 20 Vypočítajte plošný obsah časti roviny, ohraničenej grafmi funkcií $f : y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ a $g : y = 0$. Graf funkcie f sa nazýva Gaussova krivka.



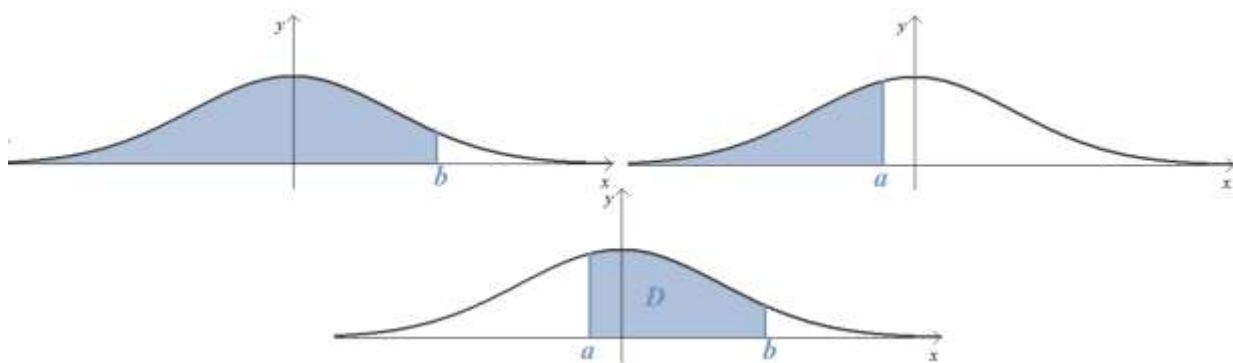
Obr. 25 Plocha ohraničená Gaussovou krivkou a x-ovou osou

Riešenie.

$$P = \int_{-\infty}^{\infty} [f(x) - g(x)] dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \left| \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi} \right| = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1$$

Poznámka. Integrál $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$ nazývame Laplaceov integrál.

Príklad 21 Ak poznáte $\int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ pre $\forall t \in \mathbb{R}$, vypočítajte plošný obsah oblasti ohraničenej krivkami $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$; $y = 0$; $x = a$; $x = b$ ($a, b \in \mathbb{R}$; $a < b$).



Obr. 26 K výpočtu určitého integrálu

$$\text{Riešenie. } P(D) = \int_{-\infty}^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx - \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Binomická veta. Pre ľubovoľné reálne čísla a, b a prirodzené číslo n platí

$$(a + b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1} b^1 + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{k} a^{n-k} b^k + \dots + \binom{n}{n-1} a^1 b^{n-1} + \binom{n}{n} a^0 b^n,$$

kde kombinačné čísla $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ sa nazývajú binomické koeficienty a $(k + 1)$ -vý člen sa rovná $\binom{n}{k} a^{n-k} b^k$. Symbolom $n!$ označujeme súčin všetkých prirodzených čísel od 1 do n , t.j. $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1) \cdot n$ a pre $n = 0$ definujeme $0! = 1$.

V Exceli pre výpočet kombinačného čísla $\binom{n}{k}$ použijeme funkciu =COMBIN(n ; k) a pre výpočet faktoriálu funkciu =FACT(n).

Príklad 22 Vypočítajte $5!$ a kombinačné číslo $\binom{7}{3}$.

Riešenie. $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = \text{FACT}(5) = 120$, $\binom{7}{3} = \frac{7!}{(7-3)!3!} = \frac{7!}{4!3!} = \text{COMBIN}(7; 3) = 35$

Výpočet v Exceli môžeme vidieť na nasledujúcom obr. 27.

	A	B
1	=FACT(5)	
2	=COMBIN(7;3)	

	A
1	120
2	35

Obr. 27 Výpočet v Exceli

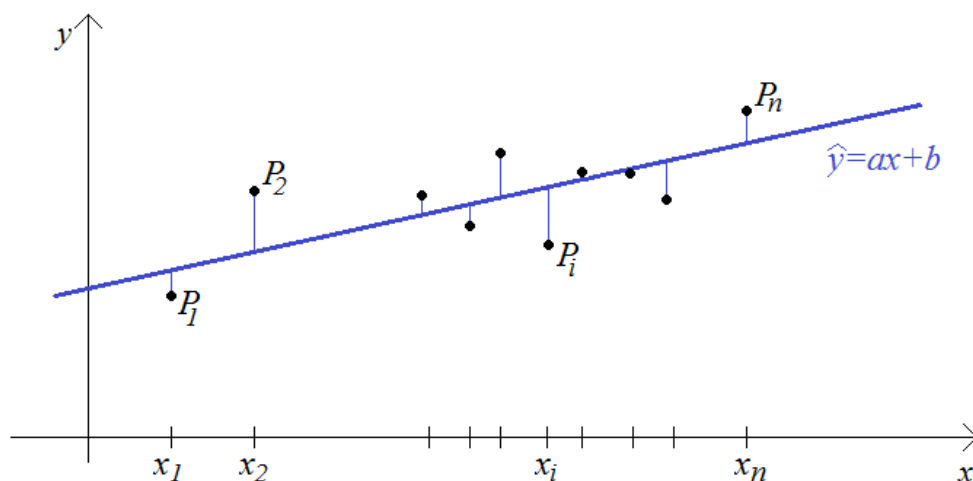
PRÍLOHA 3

Metóda najmenších štvorcov

Má široké použitie najmä v štatistike, ale aj v rôznych technických a ekonomických disciplínach.

Predpokladajme, že v rovine E_2 je daných n ($n \in N$) bodov $P_i = [x_i, y_i]$ $i = 1, 2, \dots, n$, ktoré môžu byť napr. grafickým znázornením výsledkov merania nejakej veličiny y v závislosti od hodnôt nejakej veličiny x .

Predpokladáme, že žiadne dva body P_i neležia na priamke rovnobežnej s y -ovou osou, t.j. že výsledku jedného merania x_i zodpovedá len jedna nameraná hodnota y_i . Chceme body P_i vyrovnať „čiarou“, t.j. chceme nájsť takú čiaru – graf funkcie $\hat{y} = f(x)$, ktorá by najlepšie zodpovedala grafickému priebehu bodov P_i .



Obr. 30 Podstata metódy najmenších štvorcov

Jeden z možných spôsobov takéhoto vyrovnania je, aby výraz $\sum_{i=1}^n (\hat{y}_i - y_i)^2$ nadobudol minimálnu hodnotu. Táto metóda sa nazýva metóda najmenších štvorcov. Nech sa vyrovnanie realizuje pomocou lineárnej funkcie $\hat{y} = f(x) = ax + b$, t.j. priamkou. Potom podľa metódy najmenších štvorcov má byť minimálna funkcia $F: F(a, b) = \sum_{i=1}^n (ax_i + b - y_i)^2$.

Ukážme, že F má ostré lokálne minimum. Podľa nutnej podmienky existencie lokálneho extrému funkcie dvoch premenných nájdeme parciálne derivácie prvého rádu, ktoré položíme rovné nule, aby sme našli stacionárne body funkcie F :

$$\begin{aligned}\frac{\partial F}{\partial a} &= 2 \sum_{i=1}^n (ax_i + b - y_i) \cdot x_i = 0 \\ \frac{\partial F}{\partial b} &= \sum_{i=1}^n 2(ax_i + b - y_i) = 0\end{aligned}$$

Po úprave

$$\begin{aligned} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i &= \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i + nb &= \sum_{i=1}^n y_i \end{aligned}$$

Táto sústava rovníc sa nazýva sústava normálnych rovníc.

Vynásobme prvú rovnicu n , druhú rovnicu $\sum_{i=1}^n x_i$.

$$\begin{aligned} an \sum_{i=1}^n x_i^2 + nb \sum_{i=1}^n x_i &= n \sum_{i=1}^n x_i y_i \\ a \left(\sum_{i=1}^n x_i \right)^2 + nb \sum_{i=1}^n x_i &= \sum_{i=1}^n x_i \sum_{i=1}^n y_i \end{aligned}$$

Po odčítaní rovníc dostaneme

$$a \left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] = n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

Odkiaľ pre a máme $a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$. Dá sa dokázať, že výraz v menovateli je rôzny od nuly. Ak dosadíme vypočítanú hodnotu a napr. do druhej rovnice, dostaneme $nb = \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$, odkiaľ $b = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$.

Ak $a = a_0, b = b_0$ riešením je jeden stacionárny bod $[a_0, b_0]$.

Zistíme, či v tomto bode má funkcia F lokálny extrém. K tomu vypočítame druhé parciálne derivácie funkcie F a použijeme postačujúcu podmienku existencie lokálneho extrému funkcie dvoch premenných.

$$\begin{aligned} \frac{\partial^2 F}{\partial a^2} &= 2 \sum_{i=1}^n x_i^2 \\ \frac{\partial^2 F}{\partial b^2} &= 2 \sum_{i=1}^n 1 = 2n \\ \frac{\partial^2 F}{\partial a \partial b} &= 2 \sum_{i=1}^n x_i \end{aligned}$$

$$D = \begin{vmatrix} 2 \sum x_i^2 & 2 \sum x_i \\ 2 \sum x_i y_i & 2n \end{vmatrix} = 4n \sum x_i^2 - 4(\sum x_i)^2 = 4[n \sum x_i^2 - (\sum x_i)^2] > 0$$

$D(a_0, b_0) > 0 \Rightarrow$ funkcia F má v bode $[a_0, b_0]$ lokálny extrém

$\frac{\partial^2 F}{\partial a^2} > 0 \Rightarrow$ funkcia F má v danom bode ostré lokálne minimum

Hodnoty a_0, b_0 dávajú hľadané hodnoty a, b do rovnice priamky $\hat{y} = ax + b$, ktorá vyrovnáva body P_i , t.j. „vyrovnávajúca“ priamka bude $\hat{y} = a_0x + b_0$.

Metódu najmenších štvorcov využívame aj vtedy, keď body P_i sledujú nejakú inú závislosť než lineárnu, napr. lineárne lomenú, exponenciálnu, kvadratickú, atď.

Ak napr. bodom $P_i = [x_i, y_i]$ vyhovuje nejaká kvadratická závislosť $\hat{y} = ax^2 + bx + c$, tak čísla a, b, c určíme tak, že hľadáme minimum funkcie $F: F(a, b, c) = \sum_{i=1}^n (ax_i^2 + bx_i + c - y_i)^2$. V mnohých prípadoch možno takúto úlohu vhodnou substitúciou upraviť na úlohu lineárnu. Ak napr. empirickým hodnotám vyhovuje graf funkcie $f: y = ke^{cx}, k > 0$, stačí položiť $\ln y = z, \ln k = a$ a dostaneme $\ln y = \ln k e^{cx} \Rightarrow \ln y = \ln k + cx \Rightarrow z = a + cx$, kde čísla a, b určíme z metódy najmenších štvorcov pre lineárnu funkciu. Všeobecne $\hat{y} = f(x)$ sa nazýva regresná funkcia (priamka, parabola). Pre priamku $\hat{y} = ax + b$, b sa nazýva lokujúca konštanta a a sa nazýva regresný koeficient.

Príklad 23 Nasledujúca tabuľka vyjadruje tieto závislosti y od x :

x_i	0	1	2	3	4	5
y_i	-3	0	2	0	6	6

Preložte bodmi, ktoré sú určené nameranými hodnotami $[x_i, y_i]$, priamku $y = ax + b$. Použite metódu najmenších štvorcov.

Riešenie. Vyrovnané hodnoty označme $\hat{y}_i = ax_i + b$.

Hľadáme minimum funkcie $F: F(a, b) = \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \sum_{i=1}^n (ax_i + b - y_i)^2$.

Sústava normálnych rovníc bude:

$$a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i$$

$$a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i$$

Výpočtová tabuľka v Exceli je zobrazená na nasledujúcom obr. 31.

	A	B	C	D	E
1		x_i	y_i	x_i^2	$x_i y_i$
2	1.	0	-3	=B2^2	=B2*C2
3	2.	1	0	=B3^2	=B3*C3
4	3.	2	2	=B4^2	=B4*C4
5	4.	3	0	=B5^2	=B5*C5
6	5.	4	6	=B6^2	=B6*C6
7	6.	5	6	=B7^2	=B7*C7
8	Σ :	=SUM(B2:B7)	=SUM(C2:C7)	=SUM(D2:D7)	=SUM(E2:E7)

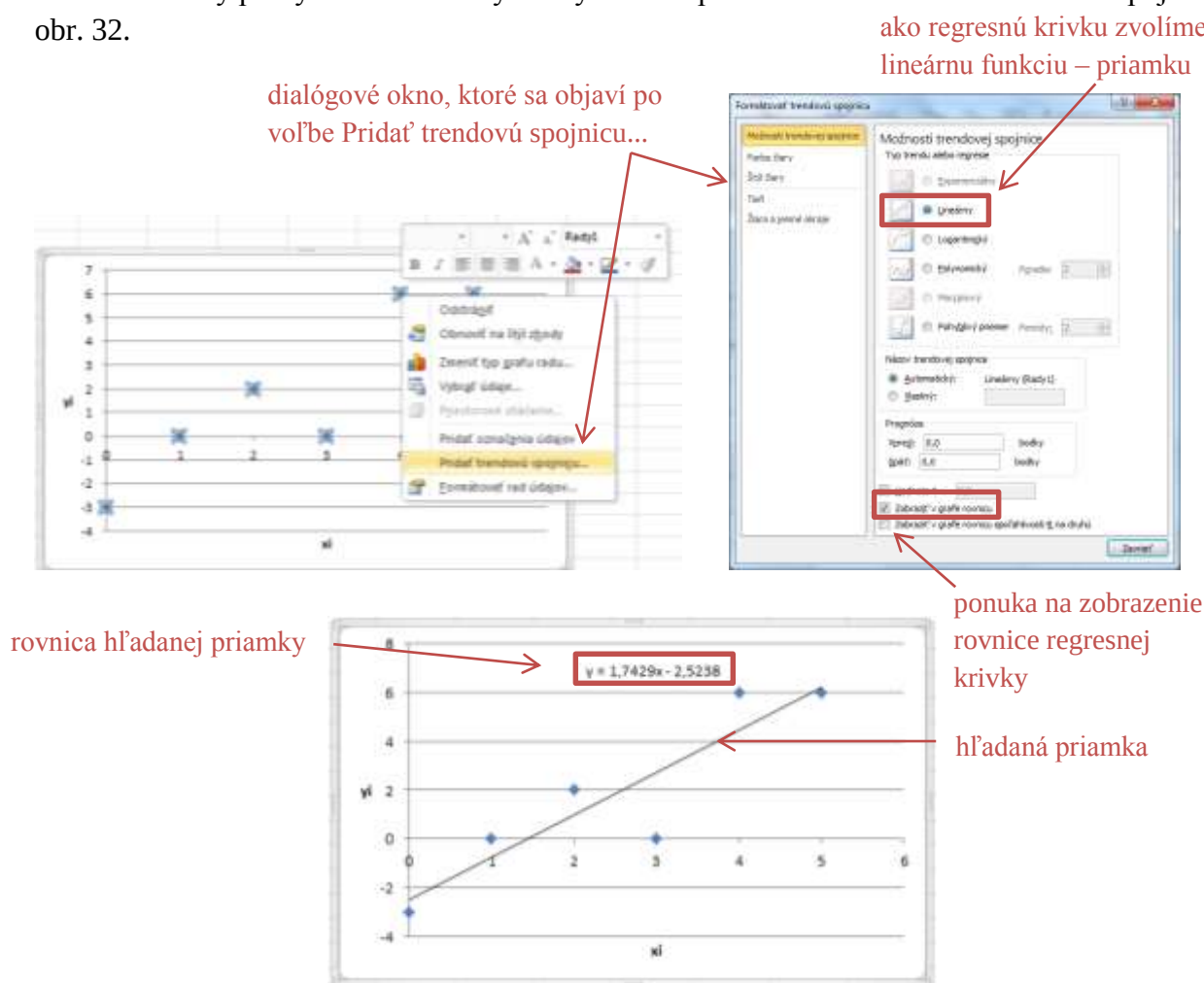
	A	B	C	D	E
1		x_i	y_i	x_i^2	$x_i y_i$
2	1.	0	-3	0	0
3	2.	1	0	1	0
4	3.	2	2	4	4
5	4.	3	0	9	0
6	5.	4	6	16	24
7	6.	5	6	25	30
8	Σ :	15	11	55	58

Obr. 31 Výpočtová tabuľka v Exceli

Z nej, po dosadení do sústavy normálnych rovníc, dostaneme

$$\begin{aligned} a \cdot 55 + b \cdot 15 &= 58 \\ a \cdot 15 + 6b &= 11 \end{aligned} \text{ , odkiaľ } a = \frac{183}{105} = 1,7429, \quad b = -\frac{265}{105} = -2,5238.$$

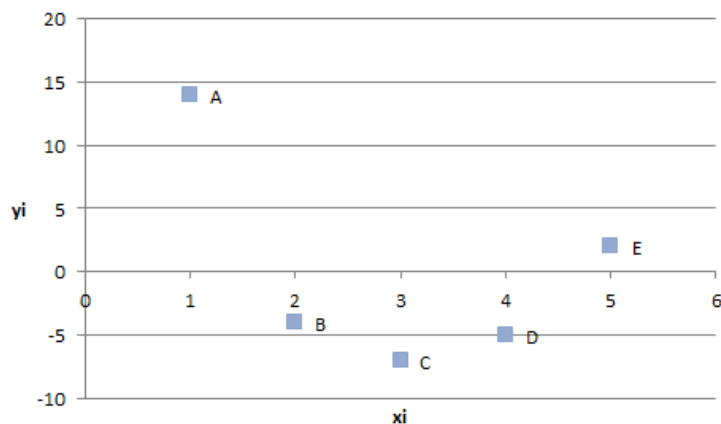
Excel umožňuje grafické riešenie zadanej úlohy, ak v bodovom grafe (XY závislosť) označíme zobrazené body pravým tlačidlom myši a vyberieme ponuku Pridať lineárnu trendovú spojniciu, obr. 32.



Obr. 32 Grafické riešenie príkladu v Exceli

Príklad 24 Bodmi $A[-3, 4]$, $B[-1, -4]$, $C[0, -7]$, $D[1, -5]$, $E[2, 2]$ metódou najmenších štvorcov preložte parabolu $y = ax^2 + b$.

Riešenie. Cez zadané body na obr. 33 chceme preložiť požadovanú krivku.



Obr. 33 Grafické zobrazenie zadania príkladu

Vyrovnané hodnoty budú ležať na grafe funkcie: $\hat{y} = ax^2 + b$.

Hľadáme minimum funkcie $F: F(a, b) = \sum_{i=1}^n (ax_i^2 + b - y_i)^2$:

$$\frac{\partial F}{\partial a} = 2 \sum_{i=1}^n (ax_i^2 + b - y_i) \cdot x_i^2 = 0$$

$$\frac{\partial F}{\partial b} = \sum_{i=1}^n 2(ax_i^2 + b - y_i) = 0$$

Sústava normálnych rovníc má tvar:

$$a \sum_{i=1}^n x_i^4 + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i^2 y_i$$

$$a \sum_{i=1}^n x_i^2 + nb = \sum_{i=1}^n y_i$$

Výpočtovú tabuľku realizujeme v Exceli, obr. 34.

	A	B	C	D	E	F
9						
10		x_i	y_i	x_i^2	x_i^4	$x_i^2 y_i$
11	A	-3	14	=B11^2	=B11^4	=B11^2*C11
12	B	-1	-4	=B12^2	=B12^4	=B12^2*C12
13	C	0	-7	=B13^2	=B13^4	=B13^2*C13
14	D	1	-5	=B14^2	=B14^4	=B14^2*C14
15	E	2	2	=B15^2	=B15^4	=B15^2*C15
16	Σ	=SUM(B11:B15)	=SUM(C11:C15)	=SUM(D11:D15)	=SUM(E11:E15)	=SUM(F11:F15)

↓

	A	B	C	D	E	F
9						
10		x_i	y_i	x_i^2	x_i^4	$x_i^2 y_i$
11	A	-3	14	9	81	126
12	B	-1	-4	1	1	-4
13	C	0	-7	0	0	0
14	D	1	-5	1	1	-5
15	E	2	2	4	16	8
16	Σ	-1	0	15	99	125

Obr. 34 Výpočtová tabuľka v Exceli

Po dosadení do sústavy normálnych rovníc dostaneme

$$a \cdot 99 + b \cdot 15 = 125, \text{ odkiaľ } a = \frac{125}{54}, b = -\frac{125}{18}.$$

Parabola má rovnicu: $y = \frac{125}{54}x^2 - \frac{125}{18}$.

Príklad 24. Predpokladajme, že máme výsledky n ($n \in \mathbb{N}$) meraní nejakej veličiny (fyzikálnej, ekonomickej...). Označme výsledky meraní $x_1, x_2, \dots, x_n$.
Nájdite hodnotu $\bar{x}$ tak, aby súčet štvorcov odchýlok nameraných hodnôt $x_1, x_2, \dots, x_n$ od čísla $\bar{x}$ bol minimálny (metódou najmenších štvorcov).

Riešenie. Táto úloha vedie k úlohe určenia lokálneho minima funkcie jednej premennej $f(x) = \sum_{i=1}^n (x_i - x)^2$. Podľa nutnej podmienky existencie lokálneho extrému funkcie nájdeme deriváciu, následne stacionárne body tejto funkcie:

$$f'(x) = 2 \sum_{i=1}^n (x_i - x) \cdot (-1) = 0 \Rightarrow \sum_{i=1}^n x_i - nx = 0 \Rightarrow x = \frac{1}{n} \sum_{i=1}^n x_i$$

Dokážeme, že v tomto stacionárnom bode má funkcia f lokálny extrém:

$f''(x) = -2 \sum_{i=1}^n (-1) = 2n > 0 \Rightarrow$ funkcia f má v danom bode lokálne minimum.
Hodnotu

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

nazývame aritmetickým priemerom hodnôt (čísel) $x_1, x_2, \dots, x_n$.

ZOZNAM BIBLIOGRAFICKÝCH ODKAZOV

1. ANDĚL, J. *Základy matematické statistiky*. Praha: MATFYZPRESS, vydavatelství Matematicko-fyzikální fakulty Univerzity Karlovy v Praze, 2011. ISBN 978-80-7378-162-0
2. BAKYTOVÁ, H., UGRON, M. *Priklady zo štatistických metód*. Bratislava: Alfa, 1976.
3. BARNET, R. A., ZIEGLER, M. R. *Finite mathematics for management, life and social sciences*. ISBN 0-02-306391-2
4. BERENSON, L.M., LEVINE, M.D. *Basic Business Statistics. Concepts and Applications. Seventh Edition*. Printed in USA, 1999. ISBN 0-13-795618-5
5. BROŽ, M., BEZVODA, V. *Excel 2007/2010. Vzorke, funkce, výpočty. Statistické zpracování dat*. Brno: Computer Press, 2011. ISBN 978-80-251-3267-8
6. BUDÍKOVÁ, M., KRÁLOVÁ, M. – MAROŠ, B. *Průvodce základními statistickými metodami*. Praha: Grada Publishing, 2010. ISBN 978-80-247-3243-5
7. CYHELSKÝ, L., HUSTOPECKÝ, J. – ZÁVODSKÝ, P. *Příklady k základům statistiky*. Praha: SNTL/ALFA, 1988.
8. CYHELSKÝ, L. *Statistika v příkladech*. Praha: SNTL, 1967.
9. GIBILISCO, S. *Statistika bez předchozích znalostí. Průvodce pro samouky*. Brno: Computer Press, 2012. ISBN 978-80-251-2465-9
10. HANOUSEK, J., CHARAMZA, P. *Moderní metody zpracování dat – matematická statistika pro každého*. Praha: Grada, 1992. ISBN 80-85623-31-5
11. HENDL, J. *Přehled statistických metod: analýza a metaanalýza dat. Třetí, přepracované vydání*. Praha: Portál, 2009. ISBN 978-80-7367-482-3
12. HRACH, K. *Sbírka úloh ze statistiky*. FSE UJEP, 2006. ISBN:80-7044-845-8
13. <http://www.mathsisfun.com/data/index.html> [dostupné dňa 05.06.2013]
14. CHAJDIK J., KOMORNÍK, J., KOMORNÍKOVÁ, M. *Štatistické metódy*. Bratislava: Statis, 1999. ISBN 80-85659-13-1
15. CHAJDIK J., RUBLÍKOVÁ, E., GUDÁBA M. *Štatistické metódy v praxi*. Bratislava: Statis, 1994. ISBN 80-85659-02-6
16. CHAJDIK, J. *Štatistické úlohy a ich riešenie v Exceli*. Bratislava: STATIS, 2005. ISBN 80-85659-39-5.
17. CHAJDIK, J. *Štatistika jednoducho*. Bratislava: STATIS, 2010. ISBN 978-80-85659-60-3.
18. JANIGA I. *Aplikovaná pravdepodobnosť a štatistika pre inžinierov. Štatistická analýza jedného a dvoch súborov*. STU, 2013. Bratislava: ISBN 978-80-227-4046-3
19. JUREČEKOVÁ, M., MOLNÁROVÁ, I. *Štatistika s Excelom*. Liptovský Mikuláš: Akadémia ozbrojených síl gen. M.R. Štefánika, 2005. ISBN 80-8040-257-4
20. Коноплёнок Е.И. – Хореева Н.К. – Лапусь А.П.. *Методическое указания по курсу „ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА“ для студентов заочной формы обучения. Московский государственный университет пищевых производств*. Москва: Изд-во АСВ, 2011.
21. LIKEŠ, J., MACHEK, J. *Matematika pro vysoké školy technické. Matematická*

- statistika*. Praha: SNTL, 1988.
22. MELOUN, M., MILITKÝ, J., HILL, M. *Statistická analýza vícerozměrných dat v příkladech*. Praha: ACADEMIA, 2012. ISBN 978-80-200-2071-0
 23. MUCHA, V. *Vybrané kapitoly z matematiky. Teória pravdepodobnosti*. Bratislava: Ekonóm, 2011. ISBN
 24. NAGY, I., PECHERKOVÁ, P., HOMOLOVÁ, J. *Příklady z pravdepodobnosti a statistiky k testum a na zkoušení*. Dostupné z <https://www.fd.cvut.cz/personal/nagyivan/Pravdep/Priklady1.pdf>
 25. NEUBAUER J., SEDLAČIK M., KŘÍŽ O. *Základy statistiky. Aplikace v technických a ekonomických oborech*. Praha: Grada Publishing, 2012. ISBN 978-80-247-4273-1
 26. ORVIS, W. J. *Microsoft Excel pro vědce a inženýry*. Praha: Computer Press, 1996. ISBN 80-85896-49-4
 27. PACÁKOVÁ, V. a kol. *Štatistika pre ekonómov. Zbierka príkladov*. Bratislava: IURA EDITION, 2005. ISBN 80-8078-032-3.
 28. PALUMBÍNY O. *Matematické metódy plánovania a vyhodnocovania experimentu*. Trnava: AlumniPress, 2009. ISBN 978-80-8096-095-7
 29. PECÁKOVÁ, I. *Statistika v terénnych průzkumech*. Příbram: PB tisk, 2011. ISBN 978-80-7431-039-3.
 30. RIMARČÍK, M. *Štatistika pre prax*. Vydané vlastným nákladem, 2007. ISBN 978-80-969813-1-1, <http://rimarcik.com>
 31. RIMARČÍK, M. *Základy statistiky*. Prešov: Fakulta zdravotníctva a sociálnej práce bl. P.P. Gojdiča. ISBN 80-969449-2-4
 32. Тихомиров В.Б. *Планирование и анализ эксперимента(при проведении исследования в легкой и текстильной промышленности)*. Москва: „Легкая индустрия“, 1974.
 33. TIRPÁKOVÁ, A., MARKECHOVÁ, D. *Štatistika v praxi s popisom postupu práce v programe Excel*. Nitra, 2008.
 34. TOŠENOVSKÝ J. *Plánování experimentu*. Ostrava: VŠB - Technická univerzita Ostrava, 2012. ISBN 978-80-248-2592-2
 35. TŮMOVÁ, O. *Metrologie a hodnocení procesů*. Praha: BEN – technická literatura, 2009. ISBN 978-80-7300-249-7
 36. VOŠICKÝ, Z. *Matematika – krok za krokem k maturite*. Fragment, Bratislava 2007. ISBN 978-80-8089-103-9
 37. ZVÁROVÁ, J. *Základy pro biomedicínske obory*. Praha: Univerzita Karlova, Karolinum, 2007. ISBN 978-80-7184-786-1

OBSAH

1. Kombinatorika.....	4
2. Teória pravdepodobnosti.....	15
2.1 Náhodný pokus, náhodný jav	15
2.2 Vzťahy medzi javmi	15
2.3 Pravdepodobnosť a jej vlastnosti.....	17
2.4 Opakované náhodné pokusy	20
2.4.1 Nezávislé opakované pokusy	20
2.4.2 Závislé opakované pokusy	21
3. Náhodná premenná.....	25
3.1 Zákon rozdelenia pravdepodobnosti diskkrétnej náhodnej premennej	26
3.2 Distribučná funkcia náhodnej premennej	29
3.3 Hustota (Zákon rozdelenia) spojitej náhodnej premennej	33
3.4 Číselné charakteristiky	35
3.4.1 Charakteristiky polohy	35
3.4.2 Charakteristiky variability	40
3.4.3 Všeobecné charakteristiky – momenty	42
3.4.4 Charakteristiky symetrie	45
3.4.5 Charakteristiky koncentrácie	45
3.5 Normovaná náhodná premenná	61
4. Rozdelenia náhodnej premennej	73
4.1 Niektoré rozdelenia diskkrétnej náhodnej premennej	73
4.1.1 Rovnomerné diskkrétne rozdelenie (Uniform discrete distribution).....	73
4.1.2 Alternatívne rozdelenie	74
4.1.3 Binomické rozdelenie (Binomial distribution).....	75
4.1.4 Poissonovo rozdelenie (Poisson distribution)	81
4.1.5 Hypergeometrické rozdelenie (Hypergeometric distribution)	84
4.2 Niektoré rozdelenia spojitej náhodnej premennej	90
4.2.1 Rovnomerné (spojité) rozdelenie (Uniform distribution (continuous))	90
4.2.2 Normálne (Gaussovo) rozdelenie (Normal (Gaussian) distribution)	93

4.2.3	Exponenciálne rozdelenie (exponential distribution).....	102
4.2.4	Weibullovo rozdelenie (Weibull distribution)	105
4.3	Rozdelenia funkcií náhodných premenných (odvodené rozdelenia, výberové rozdelenia).....	106
4.3.1	χ^2 rozdelenie (chí-kvadrát, Helmertovo-Pearsonovo rozdelenie) (Chi-squared distribution)	106
4.3.2	Studentovo <i>t</i> -rozdelenie (Student's <i>t</i> -distribution).....	107
4.3.3	<i>F</i> -rozdelenie alebo Fisherovo – Snedecorovo rozdelenie (Snedecor's <i>F</i> -distribution or Fisher-Snedecor distribution)	110
5.	Využitie analytických nástrojov a vybraných štatistických funkcií v Exceli.....	114
5.1	Opisná štatistika.....	114
5.2	Triedenie	123
5.2.1	Jednoduché triedenie	123
5.2.2	Intervalové triedenie.....	127
5.3	Skúmanie štatistickej závislosti dvoch kvantitívnych znakov (párová lineárna regresia) 131	
5.4	Plánovanie experimentu (Design Of Experiments - DOE)	154
Prílohy		171
Príloha 1		171
	Aplikácia symbolu súčtu	171
	Matice a determinanty	176
	Systém lineárnych algebraických rovníc.....	182
	Určitý integrál	189
	Nevlastný integrál	190
Príloha 2		194
	Aktivácia a použitie analytických nástrojov v Exceli	194
Príloha 3		195
	Metóda najmenších štvorcov.....	195
ZOZNAM BIBLIOGRAFICKÝCH ODKAZOV		201
Obsah.....		203

Autori: RNDr. Mária Tóthová, PhD., PaedDr. Janette Kotianová, PhD.,
Názov: Matematické metódy plánovania a vyhodnocovania experimentu
Mathematical Methods of Experiment Planning and Evaluation
Miesto vydania: Trnava
Vydavateľ: Materiálovotechnologická fakulta STU, Vydavateľstvo AlumniPress
Rok vydania: 2015
Vydanie: prvé
Rozsah : 204 strán (13,57 AH, počet obrázkov: 34,)
Edičné číslo: 3/AP/2015

ISBN 978-80-8096-227-2
EAN 9788080962272