

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

Fakulta chemickej a potravinárskej technológie

Evidenčné číslo: FCHPT-5415-25983

Modely dátových súborov

Bakalárska práca



ZADANIE BAKALÁRSKEJ PRÁCE

Študent: **Ján Karško**
ID študenta: 25983
Študijný program: automatizácia, informatizácia a manažment v chémii a potravinárstve
Kombinácia študijných odborov: 5.2.14. automatizácia, 5.2.52. priemyselné inžinierstvo
Vedúci práce: doc. RNDr. Štefan Varga, CSc.

Názov práce: **Modely dátových súborov**

Špecifikácia zadania:

Diskrétné aj spojité dátové súbory sa riadia určitými modelmi. Bakalárska práca by sa mala zaoberať štúdiom takých modelov, s ktorými sa najčastejšie stretávame v chémii. Tie modely často závisia od neznámych parametrov a ďalšou z úloh bakalárskej práce by boli odhady neznámych parametrov jednotlivých modelov pomocou dátových súborov.

Riešenie zadania práce od: 17. 02. 2014
Dátum odovzdania práce: 24. 05. 2014

L. S.

Ján Karško
študent

prof. Ing. Miroslav Fikar, DrSc.
vedúci pracoviska

prof. Ing. Miroslav Fikar, DrSc.
garant študijného programu

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

Fakulta chemickej a potravinárskej technológie

Evidenčné číslo: FCHPT-5415-25983

Modely dátových súborov

Bakalárska práca

Študijný program: automatizácia, informatizácia a manažment v chémii a potravinárstve

Číslo študijného odboru: 2621

Názov študijného odboru: 5.2.14. automatizácia, 5.2.52. priemyselné inžinierstvo

Školiace pracovisko: Ústav informatizácie, automatizácie a matematiky

Vedúci záverečnej práce: doc. RNDr. Štefan Varga, CSc.

Ďakujem vedúcemu bakalárskej práce Doc. RNDr. Štefanovi Vargovi, CSc. hlavne za inšpiráciu, pomoc pri získavaní vedomostí, pri modelovaní diskretných a spojitých dát a celkovo za vedenie, rady a pripomienky, ktoré mi poskytol pri vypracovávaní bakalárskej práce.

Abstrakt

Cieľom tejto práce, je poukázať na metódy matematickej štatistiky a ich využitia v reálnom živote, pomocou príkladov z rôznych vedných disciplín a rôznych oblastí ľudskej činnosti.

Stretávame sa tu s mnohými zákonitosťami pravdepodobnosti, kde výsledky opakovaných meraní nejakej veličiny za rovnakých podmienok sa veľmi často líšia. V skutočnosti nevieme čo to zapríčiňuje a preto tieto príklady nám pomôžu lepšie interpretovať danú problematiku.

Kľúčové slová: štatistika; diskkrétne; spojité

Abstract

The aim of this thesis is to point out the methods of mathematical statistics and their use in real life by giving examples from different scientific disciplines and areas of human activities.

We encounter with a lot of probability theory phenomena here, when the results of repeated measurements of the same parameter in the same conditions are varying. In fact we do not know the reason of this variability. For better interpretation and explanation of this problem we use examples shown in this work.

Key words: statistics; discrete; continuous

OBSAH

1	ÚVOD	8
2	MODELOVANIE DISKRÉTNÝCH DÁTOVÝCH SÚBOROV.....	9
2.1	BINOMICKÉ ROZDELENIE	14
2.1.1	<i>Príklad</i>	14
2.2	GEOMETRICKÉ ROZDELENIE.....	16
2.2.1	<i>Príklad</i>	17
2.3	POISSONOVE ROZDELENIE	18
2.3.1	<i>Príklad</i>	19
2.3.2	<i>Príklad</i>	20
2.3.3	<i>Príklad</i>	21
3	MODELOVANIE SPOJITÝCH DÁTOVÝCH SÚBOROV.....	22
3.1	NORMÁLNE (GAUSSOVE) ROZDELENIE	22
3.1.1	<i>Príklad</i>	23
3.1.2	<i>Príklad</i>	24
3.1.3	<i>Príklad</i>	25
3.1.4	<i>Príklad</i>	27
3.2	GAMA ROZDELENIE	29
3.2.1	<i>Príklad</i>	29
3.2.2	<i>Príklad</i>	31
4	VÝPOČTY V PROGRAME MATHEMATICA 8.0.....	32
5	ZÁVER.....	42
6	LITERATÚRA.....	43

1 Úvod

Dátové súbory sa v živote správajú rôzne. Môžu to byť súbory údajov zozbierané zo zisťovaní pri rôznych výskumoch, experimentoch, alebo z jednoduchých pozorovaní správania sa určitej udalosti vo viackrát opakovaných pokusoch vykonaných za rovnakých podmienok.

Na základe takýchto údajov sa štatistika opiera o poznatky z teórie pravdepodobnosti, pričom cieľom je vypožorovať nejaké závislosti a na ich základe vyvodiť určité racionálne závery. Táto teória má v dnešnom svete nezastupiteľné miesto. Dôvodom je jej využiteľnosť, či už to v technických vedách akou je chémia, v armáde, priemysle, bankovníctve a v množstve ďalších oblastí. Mojm cieľom je zaoberať sa zákonitosťami, ktorými sa riadia dátové súbory (výsledky meraní nejakej náhodnej veličiny) a ukázať na konkrétnych príkladoch, ako tieto zákonitosti fungujú. K tomu to účelu som využil matematický software Mathematica verzia 8.0 a Statgraphics. Pomocou nich som dokázal vypočítať, jednak koľko percent hodnôt môžeme očakávať v nejakom konkrétnom intervale a tiež niektoré číselné charakteristiky, ktoré dátový súbor lepšie interpretujú. Každá z týchto hodnôt nám viac vypovedá o meranej veličine. Často, zvlášť v chémii, keďže je to experimentálna veda a výsledkami sú reálne čísla, ktoré sa väčšinou chovajú ako spojité aj keď nie je výnimkou diskkrétne správanie, môžeme tieto poznatky využiť. V praxi to znamená, že vieme dopredu povedať s akou pravdepodobnosťou sa vyskytnú aké udalosti.

2 Modelovanie diskretných dátových súborov

Vo všeobecnosti sa dátové súbory delia na jednorozmerné a viacrozmerné. V tomto semestrálnom projekte sa zameriam hlavne na jednorozmerné dátové súbory. Jednorozmerné dátové súbory vznikajú ako výsledky meraní (pozorovaní) jednej veličiny na viacerých vzorkách alebo opakovane na jednej vzorke. Je zrejmé, že aj v prípade opakovaných meraní tej istej veličiny na jednej vzorke, dostávame rôzne výsledky. Hovoríme tomu, že tieto výsledky závisia od náhody. Zaujímavé je a hlavne tým sa budem zaoberať v tejto práci, že aj keď tie výsledky závisia od náhody, dajú sa vypočítavať určité zákonitosti, ktorými sa riadia. Z matematického hľadiska sú jednorozmerné dátové súbory hodnoty náhodných veličín (náhodných premenných). Náhodná veličina (budeme ju označovať X) je zobrazenie z množiny výsledkov experimentu (Ω) do množiny reálnych čísel (R). Čiže každému výsledku experimentu priradí reálne číslo.

$$X: \Omega \rightarrow R$$

Hodnoty náhodnej veličiny X budeme označovať $x_1, x_2, \dots, x_n$ a to je vlastne jednorozmerný dátový súbor. Náhodné veličiny delíme na diskkrétne a spojité. Náhodnú premennú X na pravdepodobnostnom priestore (Ω, Σ, P) nazývame diskrétna, ak jej obor hodnôt $X(\Omega) \subseteq R$ je spočítateľná množina. Teda také, ktorých hodnoty sú izolované body a je ich konečne alebo spočítateľne veľa. Príkladom diskkrétnej náhodnej veličiny môže byť dátový súbor $(0,0,1,2,0,3,5,3,2,0,1\dots)$. Diskkrétne dátové súbory sú dané tabuľkou rozdelenia pravdepodobnosti. V prvom riadku sú hodnoty, ktoré môžu nadobudnúť a v druhom riadku pravdepodobnosti s akými tie hodnoty nadobúdajú. Súčet všetkých pravdepodobnosti je rovný jednej ($\sum_i p_i = 1$).

x_i	x_1	x_2	$\dots$	x_n
p_i	p_1	p_2	$\dots$	p_n

Spojité náhodné veličiny sú také, ktorých hodnoty môžu byť ľubovoľné reálne čísla z nejakého vhodného intervalu. Hodnoty spojitých náhodných veličín sa už nedajú zadať pomocou takejto tabuľky (možných hodnôt je nespočítateľne veľa) a preto sa zadávajú pomocou hustoty rozdelenia pravdepodobnosti $f(x)$. Je to nezáporná integrovateľná funkcia, ktorá rozdeľuje pravdepodobnosť nie je jednotlivým bodom ale intervalom. Pravdepodobnosť, že hodnota spojitej náhodnej veličiny X je z intervalu $< a, b >$ sa rovná integrálu od a po b z hustoty $f(x)$.

$$P(a \leq X < b) = \int_a^b f(x)dx$$

Táto pravdepodobnosť sa dá odvodiť z definícií a vlastností hustoty $f(x)$ a distribučnej funkcie $F(X)$.

Nech (Ω, Σ, P) je pravdepodobnostný priestor a nech X je náhodná premenná na priestore udalostí (Ω, Σ) . Distribučnou funkciou náhodnej premennej X nazývame funkciu $F: R \rightarrow R$, ktorá je pre každé $x \in R$ definovaná vzťahom

$$F(x) = P[X < x]$$

Základné vlastnosti distribučnej funkcie náhodnej premennej: Nech F je distribučná funkcia náhodnej premennej X , potom platí:

1. $0 \leq F(x) \leq 1$ pre všetky $x \in R$,
2. F je neklesajúca a spojitá zľava,
3. $\lim_{x \rightarrow \infty} F(x) = 1$ a $\lim_{x \rightarrow -\infty} F(x) = 0$.

Uvediem ešte číselné charakteristiky náhodných veličín, v ktorých sú zhustené informácie o nich.

Číselné charakteristiky

- **Stredná hodnota** $E(x)$. Je najdôležitejšou číselnou charakteristikou. Je to charakteristika polohy. Udáva polohu, okolo ktorej sa nachádza najviac hodnôt (realizácií) náhodnej veličiny X .

Pre diskkrétne a spojité náhodné veličiny je definovaná vzťahmi:

$$E(X) = \sum_i x_i p_i \quad E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Často je potrebné počítať strednú hodnotu nejakej funkcie $g(X)$ náhodnej veličiny X . Táto je pre diskkrétne a spojité náhodné veličiny definovaná vzťahmi:

$$E(g(X)) = \sum_i g(x_i) p_i \quad E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- **Rozptyl (disperzia)** $D(x)$. Je charakteristikou rozptýlenia (variability). Čím je väčší, tým sú hodnoty náhodnej veličiny viac rozptýlené okolo strednej hodnoty. Pre diskkrétne a spojité náhodné veličiny je definovaný vzťahmi:

$$D(X) = \sum_i (x_i - E(X))^2 p_i \quad D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

Výpočtové vzorce pre rozptyl, pre diskkrétne aj spojité náhodné veličiny sú:

$$D(X) = \sum_i x_i^2 p_i - E(X)^2 \quad D(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - E(X)^2$$

- **Smerodajná odchýlka** $S(X)$. Je tiež charakteristikou rozptýlenia s podobnými vlastnosťami ako rozptyl a je definovaná ako odmocnina z rozptylu.

$$S(X) = \sqrt{D(X)}$$

- **Koeficient šikmosti (asymetrie) α_3 .** Je definovaný pomocou druhého a tretieho centrálneho momentu a je to charakteristika symetrie (asymetrie).

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}}$$

Keď je rovný nule ide o symetrické rozdelenie (hustota je symetrická podľa priamky $x = E(X)$). Keď je kladný, hustota je natiahnutá doprava a keď je záporný, doľava. Ešte uvediem vzťah pre k -ty centrálny moment $\mu_k (k = 2, 3, \dots)$ pre diskkrétne a spojité náhodné veličiny:

$$\mu_k = \sum_i (x_i - E(X))^k p_i \quad \mu_k = \int_{-\infty}^{\infty} (x - E(X))^k f(x) dx$$

Ak ($k = 2$) pre centrálny moment μ_k , potom platí:

$$\mu_k = D(X) = \sum_i (x_i - E(X))^2 p_i$$

Centrálne momenty sa často počítajú aj pomocou začiatočných momentov zo vzťahu:

$$\mu_k = \sum_{i=0}^k (-1)^i \binom{k}{i} v_1^i v_{k-i}$$

- **k -ty začiatočný moment $v_k (k = 1, 2, 3, \dots)$.** Je definovaný ako stredná hodnota funkcie náhodnej veličiny X^k a pre diskkrétne a spojité náhodné veličiny je definovaný vzťahmi:

$$v_k = \sum_i x_i^k p_i \quad v_k = \int_{-\infty}^{\infty} x^k f(x) dx$$

Prvý začiatočný moment je stredná hodnota. Začiatočné momenty sa často počítajú nie z definície, ale z momentovej vytvárajúcej funkcie $M_x(t)$, ktorá je definovaná ako stredná hodnota funkcie $g(X) = e^{tX}$ náhodnej veličiny X . Pre spojité náhodné veličiny platí, že

$$M_x(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

a pre diskkrétne

$$M_x(t) = E(e^{tX}) = \sum_k e^{tx_k} p_k$$

Tieto číselné charakteristiky môžeme počítať aj z momentovej vytvárajúcej funkcie, ktorá je definovaná vzťahom

$$M_x(t) = E(e^{tX}) = 1 + tv_1 + \frac{t^2}{2!} v_2 + \frac{t^3}{3!} v_3 + \dots$$

a keď urobíme jej derivácie v bode $t = 0$ dostaneme jej začiatkové momenty

$$M'_x(0) = [v_1 + tv_2 + \frac{t^2}{2!} v_3 + \dots]_{t=0} = v_1$$

$$M''_x(0) = [v_2 + tv_3 + \frac{t^2}{2!} v_4 + \dots]_{t=0} = v_2$$

$$M_x^{(k)}(0) = v_k \quad k = 1, 2, \dots, n$$

Momentová vytvárajúca funkcia sa využíva nielen na výpočet začiatkových momentov náhodnej veličiny, ale aj na určenie rozdelenia pravdepodobnosti náhodnej veličiny. Platí totiž, že ak dve náhodné veličiny majú rovnaké rozdelenie pravdepodobnosti, majú aj rovnaké momentové vytvárajúce funkcie a naopak, ak majú rovnaké momentové vytvárajúce funkcie, majú aj rovnaké rozdelenia pravdepodobnosti.

Keď sme už definovali diskkrétne a spojité náhodné veličiny a ich číselné charakteristiky, uvediem najskôr najvýznamnejšie diskkrétne rozdelenia a potom tiež najvýznamnejšie spojité rozdelenia pravdepodobnosti.

2.1 Binomické rozdelenie

Je také, kde daný experiment opakujeme n -krát (za tých istých podmienok), pri ktorých môže nastať jav A , pričom pravdepodobnosť nastátia udalosti A v každom opakovaní je $P(A) = p$. Diskrétna náhodná veličina X má binomické rozdelenie s parametrami n a p (n je prirodzené číslo a $p \in [0,1]$), ak nadobúda hodnoty $0, 1, \dots, n$ s pravdepodobnosťami

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, 2, \dots, n$$

Pre strednú hodnotu, disperziu (rozptyl) a koeficient asymetrie platí:

$$E(X) = np \quad D(X) = np(1 - p) \quad \alpha_3 = \frac{1 - 2p}{\sqrt{np(1 - p)}}$$

Pre $n = 1$ sa toto rozdelenie nazýva **alternatívne rozdelenie**.

Klasickými príkladmi binomického rozdelenia sú opakované hody mincou, kockou, počet zlých výrobkov pri opakovaných kontrolách (pričom každý výrobok musím vrátiť do zásielky, ide o tzv. *výber s opakovaním*) atď.

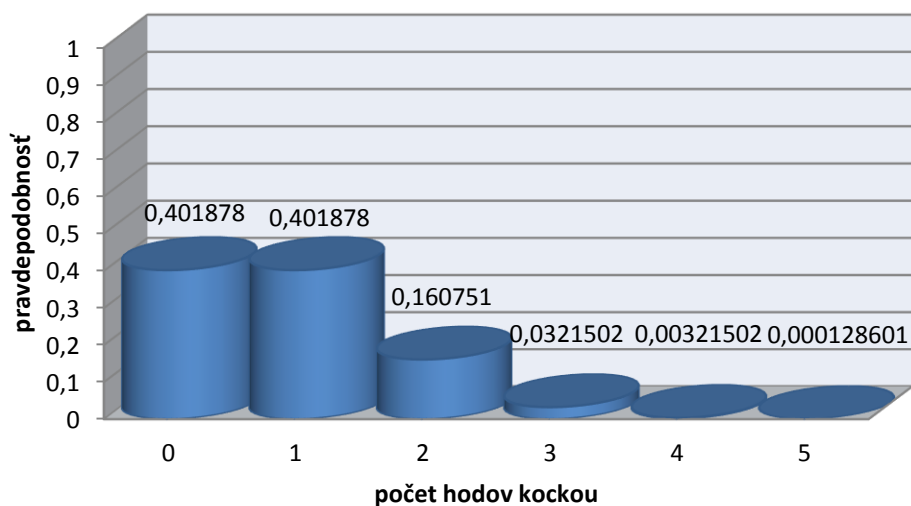
2.1.1 Príklad

Hádzem kockou 5-krát a chcem zistiť koľko krát padne 6-ka s akou pravdepodobnosťou. Určite tabuľku rozdelenia pravdepodobnosti, vypočítajte strednú hodnotu a smerodajnú odchýlku počtu padnutých šesťiek a koeficient asymetrie

Riešenie:

$$P_{n,k}(A) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Graf závislosti p od x



x_i	0	1	2	3	4	5
p_i	0,40187	0,40187	0,16075	0,032150	0,0032150	0,00012860
	8	8	1	2	2	1

Suma pravdepodobnosti $\sum p_i = 1$

Stredná hodnota $E(X) = n \cdot p = 5 \cdot \frac{1}{6} = 0,8333$

$$E(X) = \sum x_i \cdot p_i$$

Disperzia $D(x) = -n \cdot (-1 + p) \cdot p = -5 \cdot \left(-1 + \frac{1}{6}\right) \cdot \frac{1}{6} = 0,6944$

Koeficient asymetrie $\alpha_3 = \frac{1-2p}{\sqrt{np(1-n)}} = \frac{1-2 \cdot \frac{1}{6}}{\sqrt{5 \cdot \frac{1}{6} \cdot \left(1-\frac{1}{6}\right)}} = 0,8$

Pri binomickom rozdelení nás zaujíma pravdepodobnosť, že sa daná udalosť vyskytne práve pri k krát z n uskutočnených. Na základe týchto vzťahov medzi charakteristikou polohy a charakteristikou variability vždy platí táto nerovnosť:

$$E(X) > D(X)$$

Tieto charakteristiky v praxi veľmi často poznáme alebo ich vieme odhadnúť, a na ich základe vieme rozhodnúť o vhodnosti zvoleného typu rozdelenia na aproximáciu našich dát. Pre lepšie pochopenie uvediem príklad, kedy opakovane hádzem kockou 600-krát ($n = 600$). Pravdepodobnosť, že padne šestka v jednom hode je $p = \frac{1}{6}$. Stredná hodnota, že nastane udalosť A , teda padne šestka (platí však pre všetky čísla od 1 do 6) je 100. Číslo šesť by malo teoreticky padnúť 100-krát. Rozhodne však tomu nemusí byť tak. Rozptyl ukazuje ako veľmi sa môže líšiť skutočnosť od tohto teoretického predpokladu. $D(x) = -600 \cdot \left(-1 + \frac{1}{6}\right) \cdot \frac{1}{6} = 83,33$. Ak by som použil empirické pravidlo 3-sigma, potom asi 95% hodnôt by ležalo plus mínus dve smerodajné odchýlky od strednej hodnoty, v tomto prípade $100 - 2 \cdot \sqrt{D(x)}$ a $100 + 2 \cdot \sqrt{D(x)}$ tzn. v intervale $< 82; 118 >$. Ak by niektoré číslo padlo menej ako 82-krát alebo viac ako 118-krát z 600 hodov, existovala by len 5% pravdepodobnosť, že kocka je dobre vyvážená.

Výpočet v programe mathematica pozri str. 32

2.2 Geometrické rozdelenie

Experiment, ktorý má dva možné výsledky a to udalosť A s pravdepodobnosťou p a udalosť A^c s pravdepodobnosťou $(1 - p = q)$ nezávisle opakujeme až kým nenastane udalosť A . Počet opakovaní experimentu je diskrétna náhodná veličina s geometrickým rozdelením.

Diskrétna náhodná veličina X má geometrické rozdelenie pravdepodobnosti s parametrom p , ak jej hodnoty sú $k = 1, 2, 3, \dots$ s pravdepodobnosťami

$$P(X = k) = p(1 - p)^{k-1} \quad k = 1, 2, \dots$$

Pre strednú hodnotu a disperziu geometrického rozdelenia platí, že

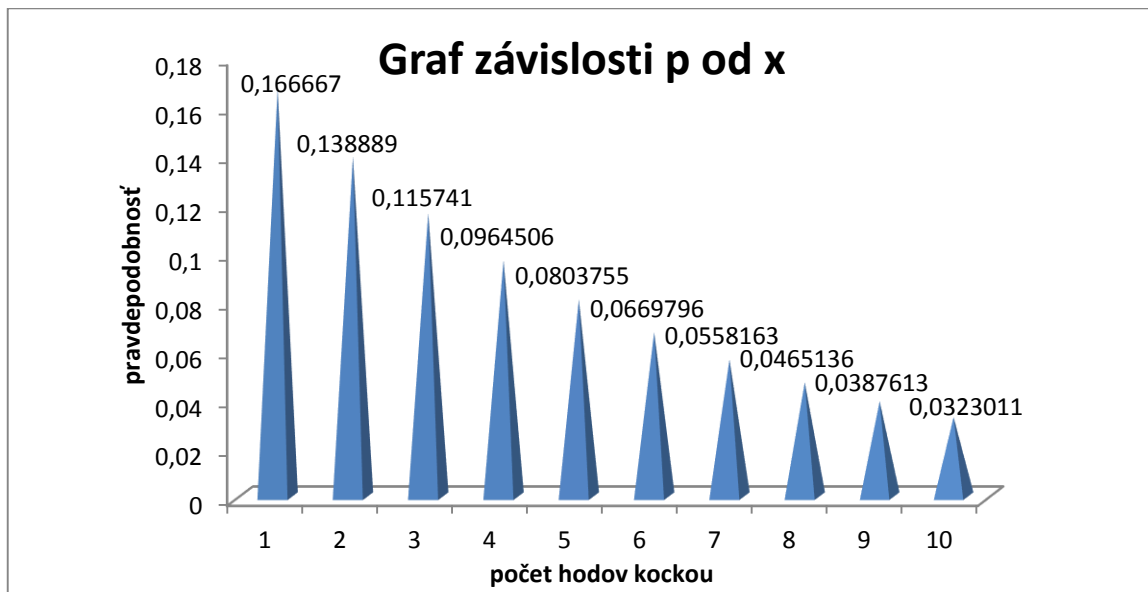
$$E(X) = \frac{1}{p} \quad D(X) = \frac{1-p}{p^2}$$

2.2.1 Príklad

Toto rozdelenie si môžeme takisto ukázať na príklade s kockou. Hádzem s kockou kým nenastane udalosť A , čiže kým nepadne 6-ka. Vypočítajte strednú hodnotu a smerodajnú odchýlku počtu hodov.

Riešenie:

$$P(X = k) = p(1 - p)^{k-1}$$



x_i	1	2	3	4	5	6	7	8	9	10	...n...
p_i	$\frac{1}{6}$	$q \cdot p$	$q^2 \cdot p$	$q^3 \cdot p$	$q^4 \cdot p$	$q^5 \cdot p$	$q^6 \cdot p$	$q^7 \cdot p$	$q^8 \cdot p$	$q^9 \cdot p$	$q^{n-1} \cdot p$

Suma pravdepodobnosti

$$\sum p_i = 1$$

$$p = \frac{1}{6} \quad q = \frac{5}{6}$$

Stredná hodnota

$$E(X) = \frac{1}{p} = \frac{1}{\frac{1}{6}} = 6$$

Disperzia

$$D(X) = \frac{1-p}{p^2} = \frac{1-\frac{1}{6}}{(\frac{1}{6})^2} = 30$$

Smerodajná odchýlka

$$S(X) = \sqrt{D(X)} = 5,477$$

Výpočet v programe mathematica pozri str. 33

2.3 Poissonove rozdelenie

Diskrétna náhodná veličina X má Poissonove rozdelenie pravdepodobnosti s parametrom $\lambda (\lambda > 0)$, ak nadobúda hodnoty $0, 1, 2, \dots$ s pravdepodobnosťami

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, 2, \dots$$

Poissonovým rozdelením sa riadi napríklad počet signálov v telefónnej ústredni, počet porúch nejakého zariadenia v konštantnom intervale, počet väd skla s konštantným plošným obsahom, počet mikroorganizmov v zornom poli mikroskopu a podobne. Pre momentovú vytvárajúcu funkciu platí, že

$$M_x(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

a odtiaľ dostaneme vzťahy pre strednú hodnotu, disperziu, koeficient asymetrie.

Vypočítame aspoň strednú hodnotu

$$E(X) = v_1 = M'_x(0) = [e^{\lambda(e^t-1)} \lambda e^t]_{t=0} = \lambda$$

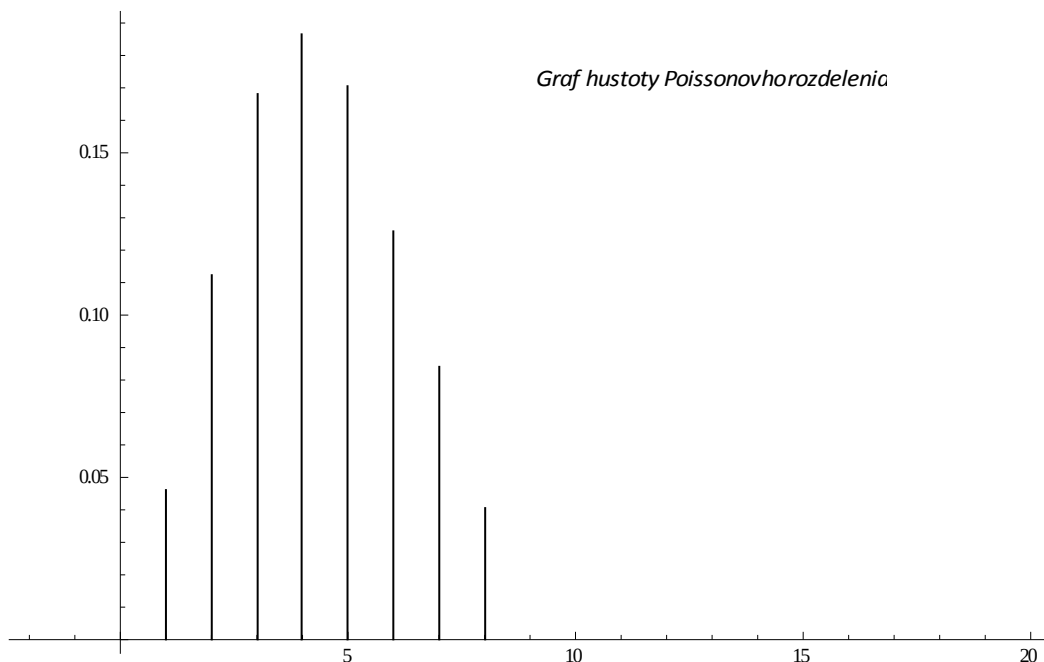
$$E(X) = D(X) = \lambda \quad \alpha_3 = \frac{1}{\sqrt{\lambda}}$$

2.3.1 Príklad

Je známe, že pri výrobe tabuľového skla určitého výrobcu sú v priemere 3 vady na 1 m². Počet vád vyrobených tabuľových skiel s obsahom 1,5 m² sú teda hodnoty, ktoré sa riadia Poissonovým rozdelením s parametrom $\lambda = 4,5$.

- Určite tabuľku rozdelenia pravdepodobnosti
- Vypočítajte koľko percent týchto skiel bude mať počet vád menší ako 3
- Vypočítajte koľko percent týchto skiel bude mať počet vád väčší ako 5

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{4,5^0 \cdot e^{-4,5}}{0!} = 0,011109$$



x_i	0	1	2	3	4	5	6	7	8
p_i	0,011	0,049	0,112	0,168	0,189	0,170	0,128	0,082	0,046
	1	9	4	7	8	8	1	3	3

$$\begin{aligned}
 P(x > 5) &= 1 - P(x \leq 5) \\
 &= 1 - 0,011109 + 0,0499905 + 0,112479 + 0,168718 \\
 &\quad + 0,189808 + 0,170827 = 0,29707
 \end{aligned}$$

Počet vád väčší ako 5 bude mať 29,707 % tabuľových skiel.

$$P(x < 3) = 0,011109 + 0,0499905 + 0,112479 = 0,17358$$

Počet vád menší ako 3 bude mať 17,358 % tabuľových skiel.

Výpočet v programe mathematica pozri str. 34

2.3.2 Príklad

Pri prenášaní binárnych dát komunikačnou linkou dochádza k zámene bitu (0 na 1 alebo 1 na 0) za 1200 μ s. Pomocou Poissonovho rozdelenia určite, aká je pravdepodobnosť, že behom 4800 μ s.

Ak za 1200 μ s dochádza k jednej zámene bitu, behom 4800 μ s dochádza k štyrom, odtiaľ dostávam priemerný počet porúch za čas, teda $\lambda = 4$.

a) Nedôjde k žiadnej chybe

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{4^0 \cdot e^{-4}}{0!} = 0.018316$$

b) Dôjde k maximálne trom chybám

$$P(x \leq 3) = e^{-4} \cdot \left(\frac{4^0}{0!} + \frac{4^1}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!} \right) = 0,43347$$

c) Dôjde k dvom až štyrom chybám

$$P(2 \leq x \leq 4) = e^{-4} \cdot \left(\frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^4}{4!} \right) = 0,53726$$

d) Dôjde k menej ako dvom chybám

$$P(x < 2) = e^{-4} \cdot \left(\frac{4^0}{0!} + \frac{4^1}{1!} \right) = 0,091578$$

Výpočet v programe mathematica pozri str. 34

Binomické rozdelenie s $n \rightarrow \infty$, s $p \rightarrow 0$ a s $np = \lambda$ sa môže aproximovať Poissonovým rozdelením s parametrom $\lambda = np$.

2.3.3 Príklad

Pravdepodobnosť toho, že 25 ročný muž prežije ďalší rok je podľa štatistických tabuliek 0,998. Poistovňa ponúka týmto poistencom, že pri ročnom poistnom vo výške 50 € vyplatí pozostalým v prípade úmrtia 10 000 €. Je poistených 1000 mužov. Aká je pravdepodobnosť, že zisk firmy ku koncu roka bude aspoň 30 000€ ? Nech X je náhodná premenná, nadobúda hodnoty $k = 0, 1, 2, \dots, 1000$ s pravdepodobnosťami:

$$P(X = k) = \binom{1000}{k} \cdot 0,002^k \cdot (1 - 0,002)^{1000-k}$$

Pre binomické rozdelenie je parameter n dosť veľké číslo a naopak parameter p malé číslo, preto možno použiť aproximáciu Poissonovým rozdelením s parametrom $\lambda = np = 1000 \cdot 0,002 = 2$.

$$P(X = 0) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{2^0 \cdot e^{-2}}{0!} = 0.135335$$

$$P(X = 1) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{2^1 \cdot e^{-2}}{1!} = 0.270671$$

$$P(X = 2) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{2^2 \cdot e^{-2}}{2!} = 0.270671$$

Hľadaná pravdepodobnosť je rovná pravdepodobnosti, že náhodná premenná X nadobudne hodnoty 0,1 alebo 2.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0,676677$$

Výpočet v programe mathematica pozri str. 35

3 Modelovanie spojitých dátových súborov

3.1 Normálne (Gaussove) rozdelenie

Gaussove rozdelenie je základným a najčastejšie sa vyskytujúcim rozdelením v spojitých dátových súboroch. Budeme ho označovať ako rozdelenie $N(\mu, \sigma)$. Kde μ je stredná hodnota a σ smerodajná odchýlka. Jeho hustota má pre $x \in R$ tvar:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

kde $\mu \in R$ a $\sigma \in R^+$ sú parametre tohto rozdelenia. Z momentovej vytvárajúcej funkcie

$$M_x(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx = e^{\mu t + \sigma^2 t^2 / 2}$$

Dostaneme jednotlivé číselné charakteristiky ako derivácie v bode 0

$$E(X) = v_1 = M'_x(0) = [e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + \sigma^2 t)]_{t=0} = \mu$$

$$v_2 = M''_x(0) = [e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + \sigma^2 t)^2 + e^{\mu t + \frac{\sigma^2 t^2}{2}} \sigma^2]_{t=0} = \mu^2 + \sigma^2$$

$$D(X) = \mu_2 = v_2 - v_1^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

Pre strednú hodnotu, smerodajnú odchýlku, koeficient asymetrie, koeficient špicatosti a že

$$E(X) = \mu \quad S(X) = \sigma \quad \alpha_3 = 0 \quad \mu_{2k-1} = 0 \quad \mu_{2k} = \frac{\sigma^{2k} (2k)!}{2^k k!}$$

Distribučná funkcia normálneho rozdelenia sa dá uviesť len v tvare integrálu

$$F(x, \mu, \sigma) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

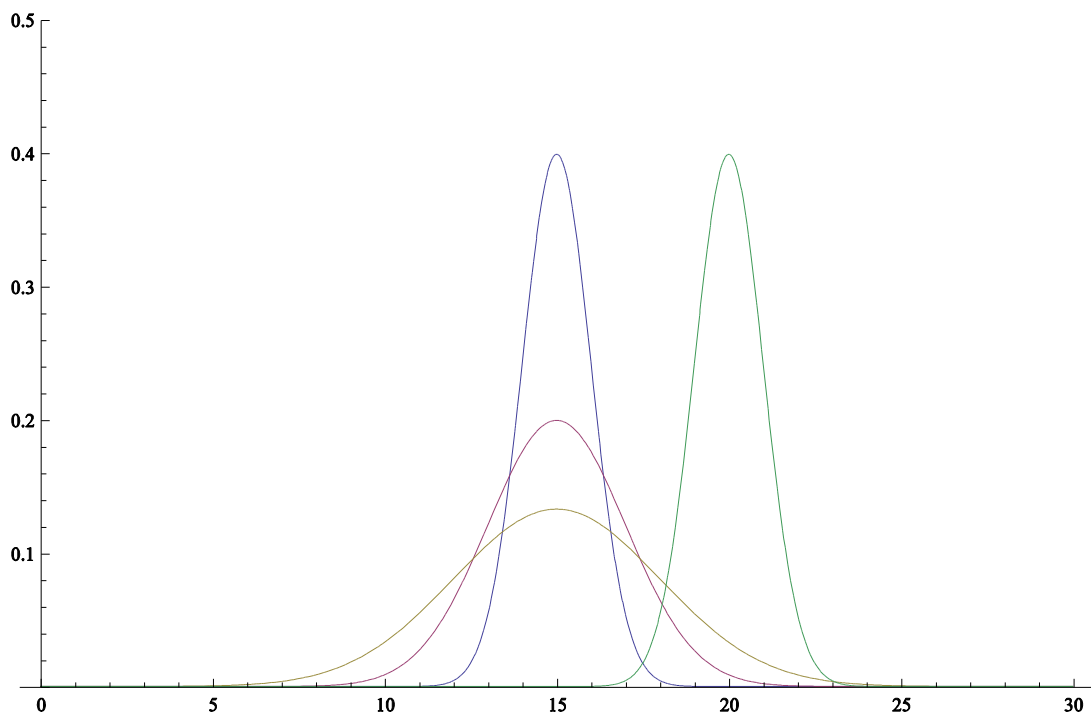
Týmto rozdelením sa riadia hodnoty opakovaných meraní, náhodné chyby meraní a rôzne deje.

3.1.1 Príklad

Nakreslite funkciu hustoty ak:

- a) $\mu = 15$; $\sigma = 1$, $\sigma = 2$, $\sigma = 3$
- b) $\mu = 20$; $\sigma = 1$

Riešenie:



Výpočet v programe mathematica pozri str. 35

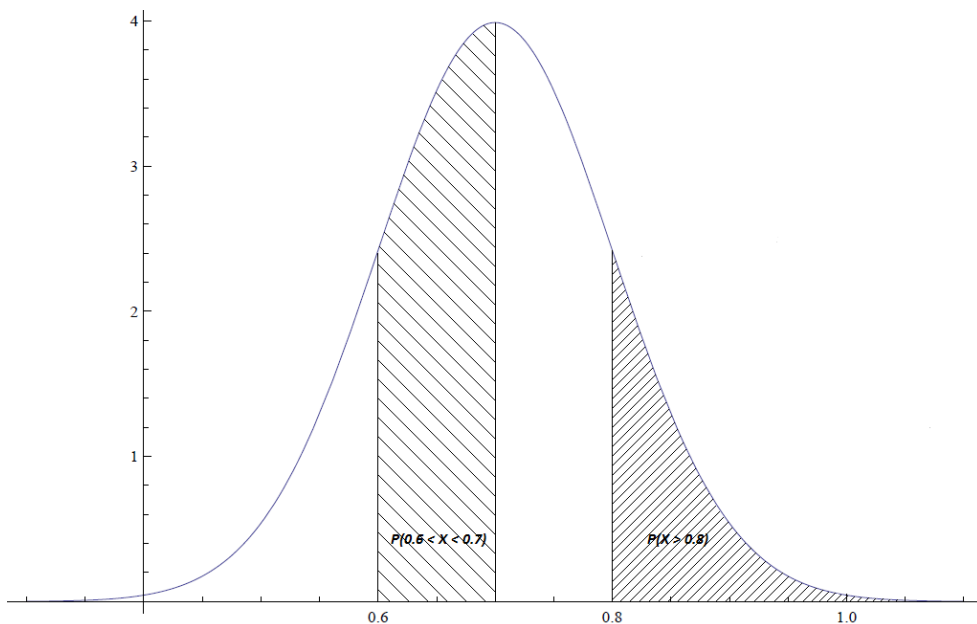
3.1.2 Príklad

Obsah železa vo vzorke sa modeluje pomocou normálneho rozdelenia so strednou hodnotou $\mu = 0,7$ a smerodajnou odchýlkou $\sigma = 0,1$.

Vypočítajte:

- Aká je pravdepodobnosť, že obsah železa vo vzorke je väčšia ako 0,8 ?
- Aká je pravdepodobnosť, že obsah železa je medzi 0,6 a 0,7 ?
- Zakreslite graf hustoty.

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



$$f_1(x) = \frac{1}{0,1\sqrt{2\pi}} e^{\frac{-(x-0,7)^2}{2 \cdot 0,1^2}}$$

- $P(X > 0,8) = \int_{0,8}^{\infty} f_1(x) dx = 0,158655$
- $P(0,6 < X < 0,7) = \int_{0,6}^{0,7} f_1(x) dx = 0,341345$

Pravdepodobnosť, že obsah železa vo vzorke je väčšia ako 0,8 je 15,8655%.

Pravdepodobnosť, že obsah železa vo vzorke je medzi 0,6 a 0,7 je 34,1345%.

Výpočet v programe mathematica pozri str. 37

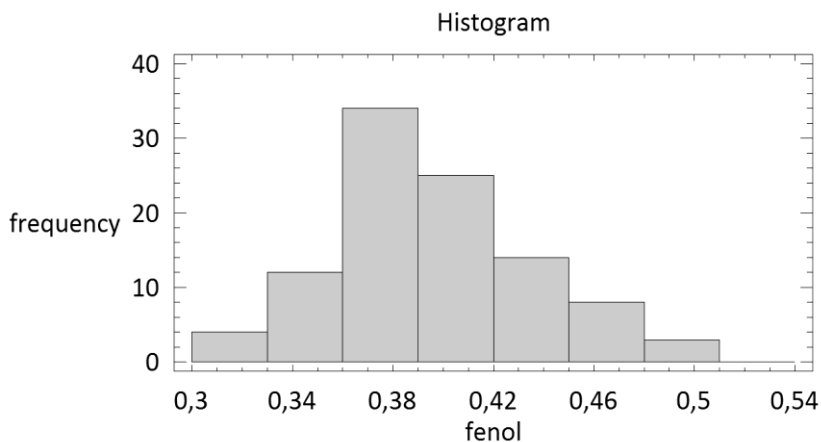
3.1.3 Príklad

Zisťovali sa hodnoty obsahu fenolu v kontrolnom roztoku, ktoré boli stanovené sériou opakovaných meraní. Ich hodnoty sú uvedené v nasledujúcej tabuľke:

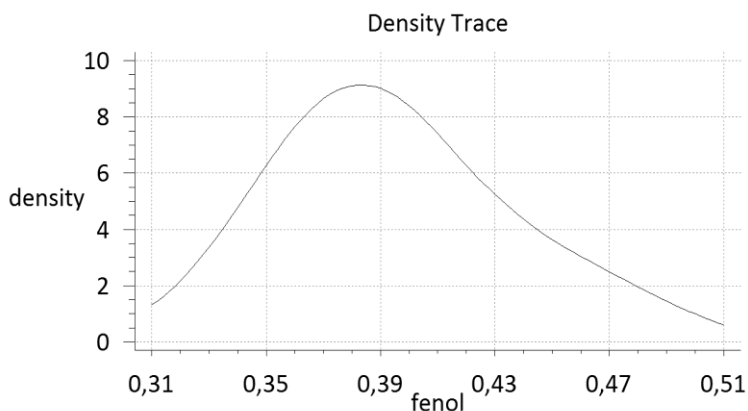
Dáta: Obsah fenolu v kontrolnom roztoku [mg/l]

0,374	0,471	0,377	0,374	0,438	0,416	0,404	0,406	0,391	0,485
0,461	0,368	0,374	0,377	0,422	0,475	0,353	0,435	0,441	0,413
0,425	0,365	0,397	0,358	0,343	0,434	0,379	0,469	0,366	0,394
0,342	0,407	0,387	0,392	0,369	0,338	0,415	0,404	0,365	0,370
0,436	0,439	0,366	0,442	0,329	0,387	0,469	0,366	0,453	0,368
0,386	0,362	0,464	0,401	0,387	0,394	0,380	0,425	0,471	0,424
0,312	0,483	0,364	0,409	0,373	0,321	0,442	0,347	0,395	0,319
0,382	0,492	0,345	0,334	0,355	0,394	0,391	0,399	0,409	0,436
0,424	0,353	0,398	0,357	0,367	0,384	0,371	0,361	0,372	0,386
0,338	0,375	0,399	0,362	0,409	0,417	0,398	0,403	0,367	0,398

Zistite akým rozdelením sa tieto dáta riadia pomocou histogramu a pomocou jadrového odhadu hustoty:



♦ **Histogram** je veľmi jednoduchý odhad hustoty $f(x)$, konštruovaný princípom, že v akom pomere sú veľkosti plôch nad jednotlivými intervalmi, v takom pomere sú aj počty hodnôt náhodného výberu v jednotlivých intervaloch.



♦ **Jadrový odhad hustoty** je podstatne presnejší odhad hustoty ako histogram. Je konštruovaný pomocou tzv. jadrovej funkcie $K(x)$. Je to funkcia definovaná na množine reálnych čísel s hodnotami v intervale $[0, \infty)$ tak, že

$$\int_{-\infty}^{\infty} K(x) dx = 1 \quad \lim_{x \rightarrow \infty} |x| K(x) = 0$$

Jadrovým odhadom hustoty $f(x)$ je

$$\hat{f}(x) = \frac{1}{n h_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right)$$

Kde $\{h_n\}_{n=1}^{\infty}$ je vhodná postupnosť kladných čísel konvergujúcich k nule.

Z výsledných grafov môžeme vyčítať, že tieto dáta by sa mali (preto, lebo sú to opakované merania) riadiť Gaussovým rozdelením. Podrobnejším správaním sa dátových súborov sa zaoberá tzv. popisná štatistika, bodovými odhadmi, testovaniami štatistickým hypotéz, korelačnou analýzou, regresnou analýzou.

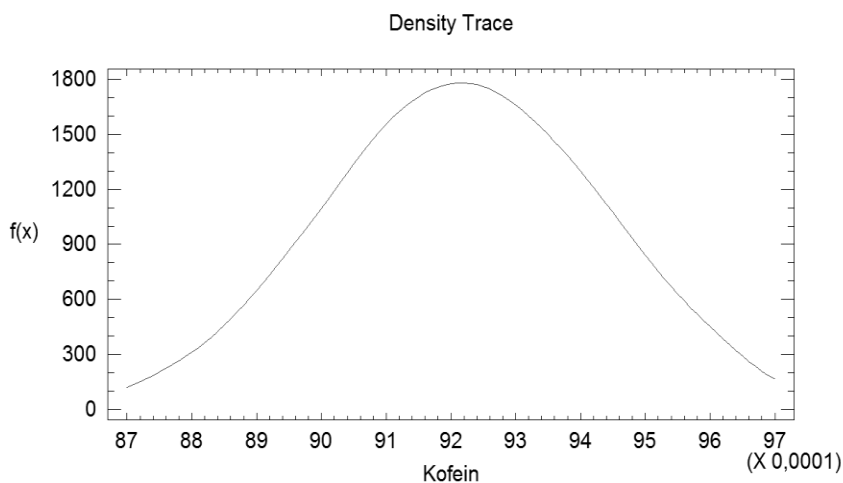
Výstup z programu Statgraphics:

Count	100
Average	0,39453
Median	0,389
Mode	0,387
Standard deviation	0,0401461
Skewness	0,45135

3.1.4 Príklad

Chromatografickou metódou sa stanovovala koncentrácia kofeínu vo vzorke opakovaným meraním. Vypočítané hodnoty sú v nasledujúcej tabuľke v $[mg/ml]$. Zistite strednú hodnotu, smerodajnú odchýlku, koeficient špicatosti a akým rozdelením sa tieto dáta riadia pomocou jadrového odhadu hustoty.

0,00876	0,00905	0,00919	0,0091	0,00926	0,0094
0,00887	0,00906	0,00921	0,00912	0,00931	0,00941
0,00887	0,00908	0,00926	0,00912	0,00941	0,00941
0,00892	0,00908	0,00927	0,00915	0,00953	0,00945
0,00893	0,0091	0,00927	0,00916	0,00953	0,00946
0,00894	0,00906	0,00927	0,00918	0,0093	0,00947
0,00894	0,00909	0,00927	0,00921	0,0093	0,00947
0,00901	0,00912	0,00928	0,00922	0,00932	0,0095
0,00904	0,00916	0,00929	0,00924	0,00938	0,00956
0,00905	0,00918	0,00929	0,00924	0,00939	0,00966



$$\mu = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0,0092195; \sigma = s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0.00019191$$

$$\alpha_3 = A_3 = \frac{M_3}{M_2^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2^{3/2}} = -0,0491424$$

Výberové číselné charakteristiky korešpondujú s vyššie definovanými číselnými charakteristikami náhodných veličín a majú obdobné vlastnosti. Výberový aritmetický priemer $\bar{x}$ je najlepším nevychýleným odhadom strednej hodnoty μ , výberová smerodajná odchýlka s je najlepším nevychýleným odhadom smerodajnej odchýlky σ , výberový koeficient šikmosti (asymetrie) A_3 definovaný podielom tretieho a druhého výberového centrálného momentu umocneného na tri polovice je najlepším nevychýleným odhadom koeficientu šikmosti (asymetrie) α_3 . Záporná hodnota koeficientu asymetrie poukazuje na mierne natiahnutie hustoty doľava.

♦ **Výberový k -ty centrálny moment** $M_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$; ($k = 2, 3, \dots$).

♦ **Najlepší nevychýlený odhad.** Odhad $\hat{\theta} = g(x_1, x_2, \dots, x_n)$ neznámeho parametra θ sa nazýva najlepším nevychýleným odhadom, ak je nevychýlený a ak pre jeho disperziu platí, že

$$D(g(x_1, x_2, \dots, x_n)) \leq D(h(x_1, x_2, \dots, x_n))$$

kde $h(x_1, x_2, \dots, x_n)$ je ľubovoľný nevychýlený odhad parametra θ .

3.2 Gama rozdelenie

Riadi sa ním hlavne životnosť výrobkov a spoľahlivosť zariadení. Hustota gama rozdelenia závisí od parametrov $m, \lambda (m > 0, \lambda > 0)$. Parameter λ môžeme nazvať poruchovým parametrom a parameter m tvarovým parametrom.

$$f(x, m, \lambda) = f(x) = \begin{cases} \lambda^m x^{m-1} e^{-\lambda x} / \Gamma(m), & x \geq 0 \\ 0, & x < 0 \end{cases}$$

a pre momentovú vytvárajúcu funkciu dostaneme, že

$$M_x(t) = \frac{\lambda^m}{\Gamma(m)} \int_0^{\infty} x^{m-1} e^{tx-\lambda x} dx = (1 - \lambda^{-1}t)^{-m}$$

Pre najvýznamnejšie číselné charakteristiky platí, že

$$E(X) = m\lambda^{-1} \quad D(X) = m\lambda^{-2} \quad \alpha_3 = \frac{2}{\sqrt{m}} \quad \alpha_4 = \frac{6}{m}$$

Treba ešte uviesť, že $\Gamma(m)$ je tzv. gama funkcia, definovaná ako nevlastný integrál vzťahom

$$\Gamma(m) = \int_0^{\infty} x^{m-1} e^{-x} dx$$

Pre prirodzené čísla platí, že $\Gamma(m) = (n-1)!$.

3.2.1 Príklad

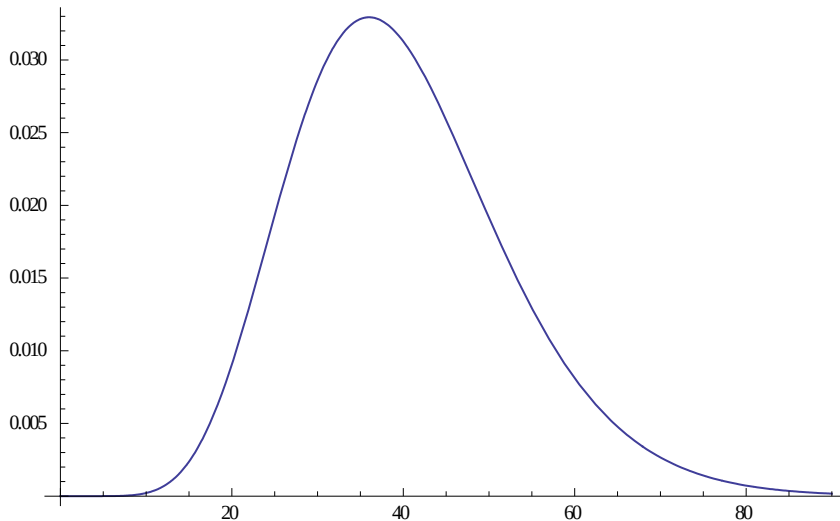
Životnosť pozinkovaných plechov je náhodná veličina, ktorá sa riadi Gama rozdelením so strednou hodnotou $E(X) = 40$ a parametrami $m = 10, \lambda = 0,25$.

Vypočítajte:

- Koľko percent plechov prežije 40 rokov ?
- Koľko percent plechov sa poškodí za 20 rokov ?

- c) Koľko percent plechov sa poškodí medzi 10. a 20. rokom ?
- d) Vypočítajte koeficient šikmosti pomocou centrálnych momentov.

Riešenie:



Pozn. Začína v nule a je natiahnuté doprava.

$$f_2(x) = \frac{\lambda^m x^{m-1} e^{-\lambda x}}{\Gamma(10)}$$

a) $P(X > 40) = \int_{40}^{\infty} f_2(x) dx = 0,54207$

b) $P(X < 20) = \int_0^{20} f_2(x) dx = 0,03183$

c) $P(10 < X < 15) = \int_{10}^{15} f_2(x) dx = 0,005$

d) $\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\int_{-\infty}^{\infty} (x-E(X))^3 f_2(x) dx}{\int_{-\infty}^{\infty} (x-E(X))^2 f_2(x) dx^{3/2}} = \frac{1280}{160^{3/2}} = 0,632456$

Pravdepodobnosť pozinkovaných plechov, že neskorodujú 40 rokov je 54,21%.

Pravdepodobnosť korózie pozinkovaných plechov do 20 rokov je 3,18%.

Medzi 10. a 15. rokom je pravdepodobnosť skorodovania pozinkovaných plechov 0,5%. Koeficient šikmosti je kladný, hustota je natiahnutá doprava.

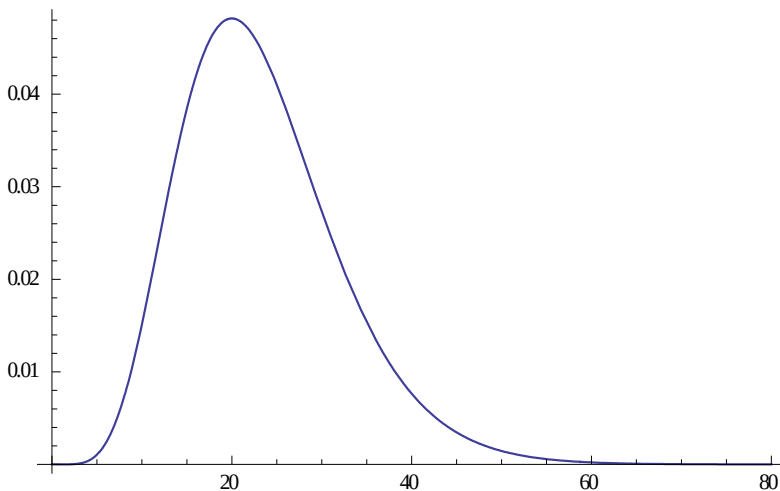
Výpočet v programe mathematica pozri str. 38

3.2.2 Príklad

Životnosť (v mesiacoch) určitého druhu výrobkov je náhodná veličina, ktorá sa riadi gama rozdelením s parametrami $m = 7, \lambda = 0,3$.

- Určite hustotu rozdelenia pravdepodobnosti a nakreslite jej graf
- Vypočítajte koľko percent výrobkov tohto druhu prežije 30 mesiacov
- Vypočítajte koľko percent výrobkov tohto druhu sa pokazí do 20 mesiacov
- Vypočítajte strednú hodnotu životnosti týchto výrobkov

$$f_3(x) = \frac{\lambda^m x^{m-1} e^{-\lambda x}}{\Gamma(7)}$$



$$P(X > 30) = \int_{30}^{\infty} f_3(x) dx = 0.206785$$

$$P(X < 20) = \int_0^{20} f_3(x) dx = 0.393697$$

$$E(X) = m\lambda^{-1} = \frac{7}{0,3} = 23.33$$

Výpočet v programe mathematica pozri str. 40

4 VÝPOČTY V PROGRAME MATHEMATICA 8.0

$\text{Bi}[p,n,k]=\text{Binomial}[n,k]*p^k*(1-p)^{(n-k)}$

$(1-p)^{-k+n} p^k \text{Binomial}[n,k]$

$\text{Sum}[\text{Bi}[p,n,k],\{k,0,n\}]$

1

$\text{Sum}[\text{Exp}[t*k]*\text{Bi}[p,n,k],\{k,0,n\}]$

$(1+(-1+e^t) p)^n$

$\text{Mx}[t_]= (1-p+p*\text{Exp}[t])^n$

$(1-p+e^t p)^n$

$N1=\text{Mx}'[0]$

$n p$

$N2=\text{Mx}''[0]$

$n p+(-1+n) n p^2$

$M2=N2-N1^2$

$n p+(-1+n) n p^2-n^2 p^2$

$\text{Simplify}[\%]$

$-n (-1+p) p$

$\text{Simplify}[\%]$

$\text{Bi}[0.166667,5.,5.]$

$\text{Bi}[1/6,5,5]$

$\text{Bi}[1/6,5,5]$

$\text{Bi}[1/6,5,5]/N$

$\text{Bi}[0.166667,5.,5.]$

$N[\text{Table}[\text{Bi}[1/6,5,k],\{k,0,5,1\}]]$

$\{\text{Bi}[0.166667,5.,0.],\text{Bi}[0.166667,5.,1.],\text{Bi}[0.166667,5.,2.],\text{Bi}[0.166667,5.,3.],\text{Bi}[0.166667,5.,4.],\text{Bi}[0.166667,5.,5.]\}$

$\text{Sum}[\text{Bi}[1/6,5,k],\{k,0,5\}]$

$Bi[1/6,5,0]+Bi[1/6,5,1]+Bi[1/6,5,2]+Bi[1/6,5,3]+Bi[1/6,5,4]+Bi[1/6,5,5]$

$Sum[Bi[1/6,5,k],\{k,0,5\}]/N$

$Bi[0.166667,5.,0.]+Bi[0.166667,5.,1.]+Bi[0.166667,5.,2.]+Bi[0.166667,5.,3.]+$

$Bi[0.166667,5.,4.]+Bi[0.166667,5.,5.]$

[Spät na zadanie príkladu](#)

$Geom[p_,k_]=p*(1-p)^(k-1)$

$(1-p)^{-1+k} p$

$Sum[Exp[t*k]*Geom[p,k],\{k,1,\infty\}]$

$(e t p)/(1-e t+e t p)$

$Mx[t_]=p*Exp[t]/(1-Exp[t]+p*Exp[t])$

$(e t p)/(1-e t+e t p)$

$N1=Simplify[Mx'[0]]$

$1/p$

$N2=Simplify[Mx''[0]]$

$(2-p)/p^2$

$N3=Simplify[Mx'''[0]]$

$(6-6 p+p^2)/p^3$

$M2=Simplify[N2-N1^2]$

$(1-p)/p^2$

$M3=Simplify[N3-3*N1*N2+2*N1^3]$

$(2-3 p+p^2)/p^3$

$E(x)=1-p/p=(1-(1/6))/(1/6)$

Set::write: Tag Plus in -1+1 is Protected.

Set::write: Tag Times in e x is Protected.

5

```
Table[N[Geom[1/6,k]],{k,1,10,1}]
{0.166667,0.138889,0.115741,0.0964506,0.0803755,0.0669796,0.0558163,0.0465136,0.0387613,0.0323011}
```

$q=5/6$

$5/6$

$p=1/6$

$1/6$

$\text{Sum}[q^{(n-1)}p,\{n,1,\infty\}]$

1

[Späť na zadanie príkladu](#)

$\text{Po}[\lambda,k]=\frac{\lambda^k \text{Exp}[-\lambda]}{k!}$

$(E^{-\lambda} \lambda^k)/k!$

$\text{Sum}[\text{Po}[\lambda,k],\{k,0,\infty\}]$

1

$\text{Table}[N[\text{Po}[4.5,k]],\{k,0,8,1\}]$

{0.011109,0.0499905,0.112479,0.168718,0.189808,0.170827,0.12812,0.0823629,0.0463292}

$\text{Sum}[\text{Po}[4.5,k],\{k,6,\infty\}]$

0.29707

$\text{Sum}[\text{Po}[4.5,k],\{k,0,2\}]$

0.173578

[Späť na zadanie príkladu](#)

$\text{Po}[\lambda,k]=\frac{\lambda^k \text{Exp}[-\lambda]}{k!}$

$(E^{-\lambda} \lambda^k)/k!$

$\text{Sum}[\text{Po}[\lambda,k],\{k,0,\infty\}]$

1

$\text{Table}[N[\text{Po}[4,k]],\{k,0,7,1\}]$

{0.0183156,0.0732626,0.146525,0.195367,0.195367,0.156293,0.104196,0.0595404}

Sum[Po[4,k],{k,0,0}]

1/E^4

Sum[Po[4,k],{k,0,3}]

71/(3 E^4)

Sum[Po[4,k],{k,2,4}]

88/(3 E^4)

Sum[Po[4,k],{k,0,1}]

5/E^4

[Späť na zadanie príkladu](#)

Po[\[Lambda]_, k_] = (\[Lambda]^k*Exp[-\[Lambda]])/k!

(E^-\[Lambda] \[Lambda]^k)/k!

Sum[Po[\[Lambda], k], {k, 0, \[Infinity]}]

1

Sum[Po[2, k], {k, 0, 2}] // N

0.676676

[Späť na zadanie príkladu](#)

Clear[f1,f2,f3,f4]

f1[x_]=Exp[-(x-15)^2/2]/(Sqrt[2*Pi]*1)

$$e^{-\frac{1}{2}((x-15) \times 2)} / \sqrt{2}$$

f2[x_]=Exp[-(x-15)^2/8]/(Sqrt[2*Pi]*2)

$$e^{-\frac{1}{8}((x-15) \times 2)} / (2 \sqrt{2})$$

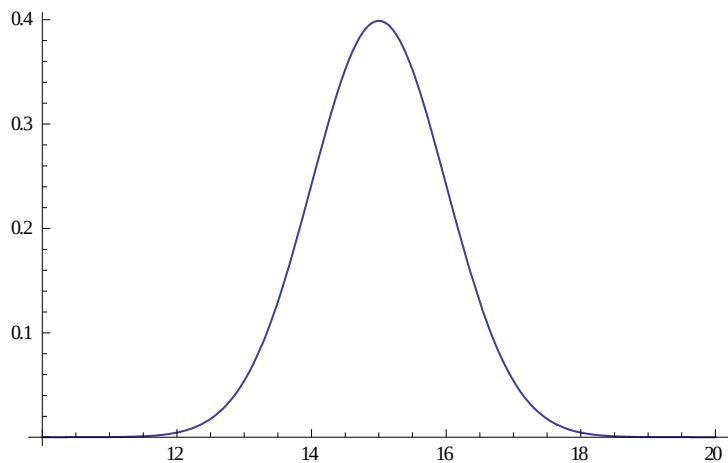
f3[x_]=Exp[-(x-15)^2/18]/(Sqrt[2*Pi]*3)

$$f_1(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-15)^2}{2}\right)$$

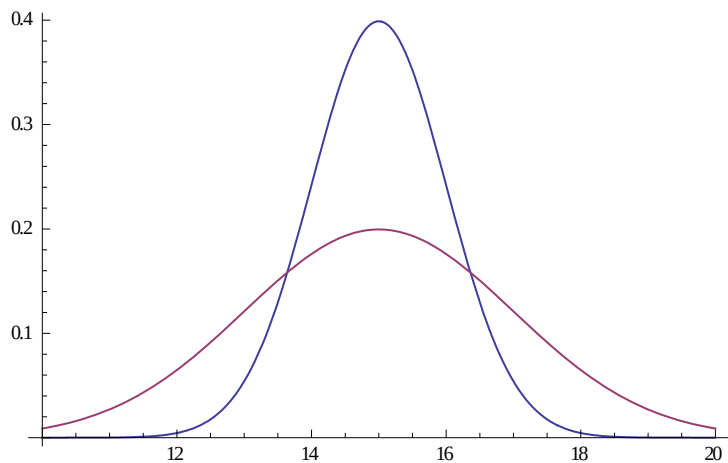
$$f_4(x) = \frac{\exp\left(-\frac{(x-20)^2}{2}\right)}{\sqrt{2\pi}}$$

$$f_2(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-20)^2}{2}\right)$$

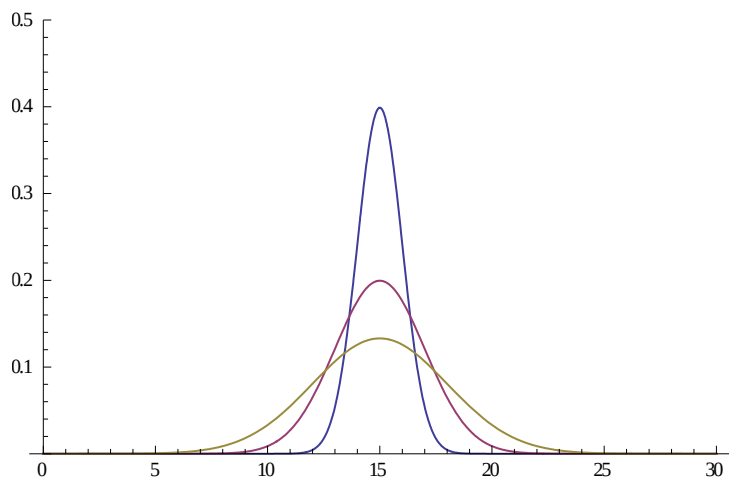
Plot[f1[x],{x,10,20}]



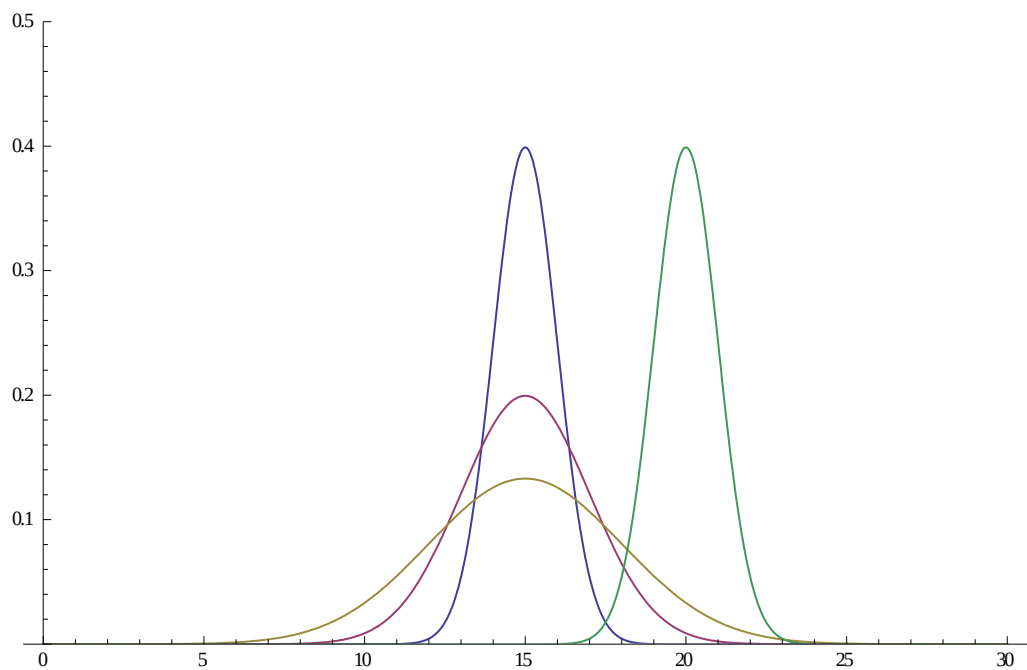
Plot[{f1[x],f2[x]},{x,10,20}]



Plot[{f1[x],f2[x],f3[x]},{x,0,30},PlotRange->{0,0.5}]



`Plot[{f1[x], f2[x], f3[x], f4[x]}, {x, 0, 30}, PlotRange -> {0, 0.5}]`



[Spät na zadanie príkladu](#)

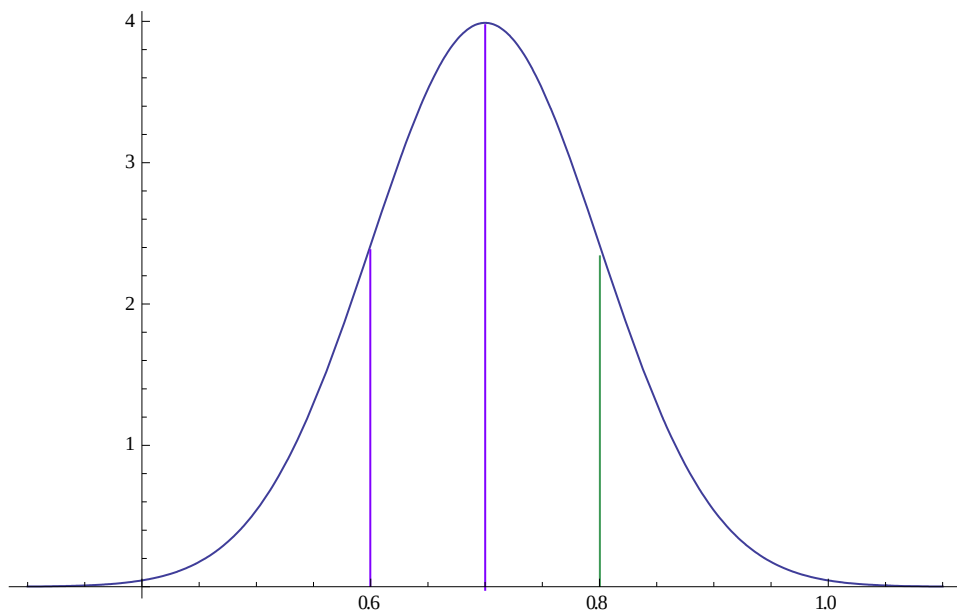
`f1[x_]=Exp[-(x-0.7)^2/0.02]/(Sqrt[2*Pi]*0.1)`

`3.98942 Exp[-50. (x-0.7)^2]`

```
Integrate[f1[x],{x,-∞,∞}]
```

```
1._6.00077×10-15 ĩ
```

```
Plot[f1[x],{x,0.3,1.1}]
```



```
Integrate[f1[x],{x,0.6,0.7}]/N
```

```
0.341345
```

```
Integrate[f1[x],{x,0.8,∞}]/N
```

```
0.158655_6.00077×10-15 ĩ
```

[Späť na zadanie príkladu](#)

```
Clear[f1]
```

```
f1[x_]=0.25^10*x^9*Exp[-0.25*x]/Gamma[10]
```

```
2.62807×10-12 e-0.25 x x9
```

```
Integrate[f1[x],{x,0,40}]
```

```
0.54207
```

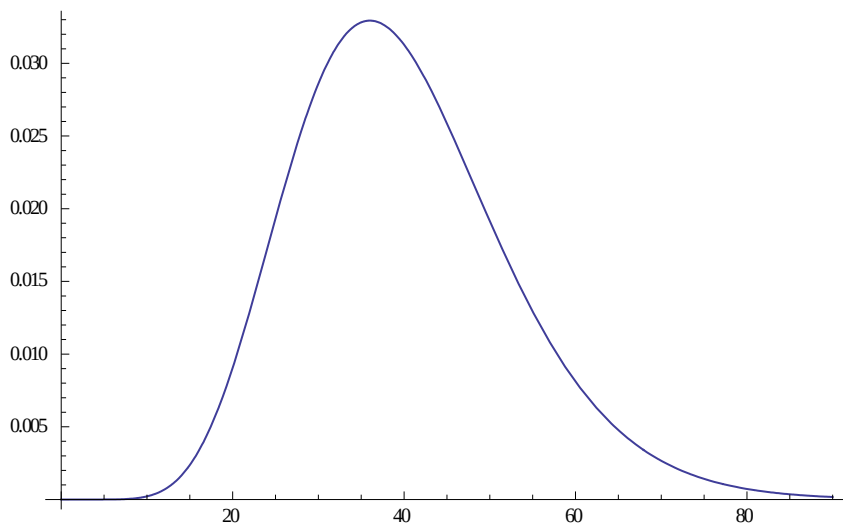
```
Integrate[f1[x],{x,0,20}]
```

```
0.0318281
```

```
Integrate[f1[x],{x,10,15}]
```

```
0.00502982
```

Plot[f1[x],{x,0,90}]



e=40

40

l=0.25

0.25

m=10

10

rm=\!(\

*SubsuperscriptBox[\(\([Integral]\), \((0)\), \(\([Infinity]\))\)](x^\((m - 1)\))*E^\((\(-x\))\)) \[DifferentialD]x\)\)

362880

f2[x_]=((l^m)*(x^(m-1))*(E^(-l*x)))/rm//N

2.62807*10^-12 2.71828^(-0.25 x) x^9

c2=(x-e)^2*f2[x]

2.62807*10^-12 2.71828^(-0.25 x) (-40+x)^2 x^9

Integrate[c2,{x,0,\[Infinity]}}

160.

c3=(x-e)^3*f2[x]

$$2.62807 \cdot 10^{-12} \cdot 2.71828^{(-0.25 x)} \cdot (-40+x)^3 \cdot x^9$$

$$\text{Integrate}[c3, \{x, 0, \infty\}]$$

$$1280.$$

$$\text{alfa3} = 1280 / 160^{(3/2)} // N$$

$$0.632456$$

$$2 / \sqrt{10} // N$$

$$0.632456$$

[Späť na zadanie príkladu](#)

$$f[x_, m_, \lambda_] = \lambda^m \cdot x^{(m-1)} \cdot \text{Exp}[-\lambda \cdot x] / \text{Gamma}[m]$$

$$(\text{e}^{-x} \cdot \lambda \cdot x^{-1+m} \cdot \lambda^m) / \text{Gamma}[m]$$

$$\text{Integrate}[f[x, 7, 0.3], \{x, 0, \infty\}]$$

$$1.$$

$$f[x, 7, 0.3]$$

$$3.0375 \times 10^{-7} \cdot \text{e}^{-0.3 x} \cdot x^6$$

$$\text{Integrate}[f[x, 7, 0.3], \{x, 30, \infty\}]$$

$$0.206781$$

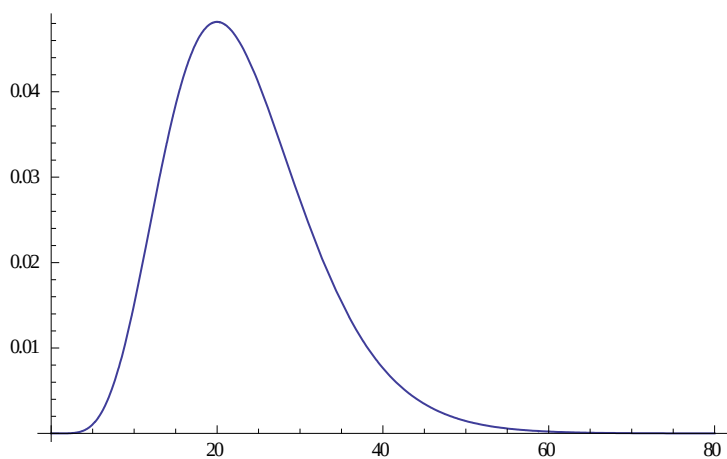
$$\text{Integrate}[f[x, 7, 0.3], \{x, 0, 20\}]$$

$$0.393697$$

$$7/0.3$$

$$23.3333$$

$$\text{Plot}[f[x, 7, 0.3], \{x, 0, 80\}]$$



[Spät na zadanie príkladu](#)

5 Záver

Náplňou tejto bakalárskej práce bolo poukázať na reálne využitie teoretických modelov, ktorými sa riadia diskkrétne alebo spojité dátové súbory. Uviedol som definície a základné vlastnosti rozdelení, ktoré sa najčastejšie využívajú v štatistických analýzach. K tomuto cieľu sa použili tri modely diskrétnych dátových súborov a dva modely spojitých dátových súborov. Stanovené úlohy sa podarilo splniť a kvôli lepšej implementácii sa riešili aj grafickým výpočtom. Zároveň mojou snahou bolo, poskytnúť pre čitateľa základné informácie z oblasti štatistickej analýzy jednorozmerných dát a nevyhnutné softvérové minimum. Na reálne využitie štatistiky v chémii chcem ešte poukázať na možnosť použitia štatistických testov dobrej zhody náhodného výberu s modelom, čo by mohlo byť rozšírením tejto práce.

6 Literatúra

[1] Milan Meloun, Jiří Milítký: Kompendium statistického zpracování dat, *Metody a řešené úlohy*, Praha, (2. vyd.) 2006, ISBN 80-200-1396-2

[2] Jozef Chajdiak, Eva Rublíková, Miroslav Gudába: Štatistické metódy v praxi, Bratislava 1994, ISBN 80-85659-02-6

[3] <http://www.kirp.chtf.stuba.sk/moodle/>, Aplikovaná štatistika, Doc. RNDr. Štefan Varga, CSc.

[4] Štefan Varga, Matematická štatistika, Bratislava STU-2012, ISBN 978-80-227-3789-0

[5] Radoslav Harman, Erika Hönschová, Ján Somorčík: Zbierka úloh zo základov teórie pravdepodobnosti, Bratislava, 2009, ISBN 978-80-89186-53-2

[6] http://www.km.fpv.ukf.sk/upload_publikacie/20120125_143707__1.pdf