

ČÍSELNÉ CHARAKTERISTIKY NÁHODNEJ PREMENNEJ X

- stručná (prehľadná) informácia o náhodnej premennej vyjadrená jedným číslom

$F(x)$, $f(x)$ resp. $p(x)$ popisujú rozdelenie náhodnej premennej jednoznačne, t.j. do všetkých podrobností. Niekedy nás zaujíma iba jeden aspekt náhodnej premennej:

očakávaná hodnota, hodnota pod ktorou leží iba určité množstvo hodnôt náhodnej premennej, ich variabilita, šikmosť rozdelenia či koncentrácia hodnôt náhodnej premennej okolo očakávanej hodnoty

Číselné charakteristiky náhodnej premennej X — 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 (koeficient šikmosti);
- charakteristiky koncentrácie (koeficient špicatosti).

Charakteristiky polohy náhodnej premennej X

Stredná hodnota – Mean – EXPECTED VALUE (očakávaná

hodnota) – $E(X)$ je číslo, ktoré opisuje, okolo ktorej hodnoty na číselnej osi sú sústredené samotné hodnoty náhodnej premennej X s ohľadom na ich pravdepodobnosti:

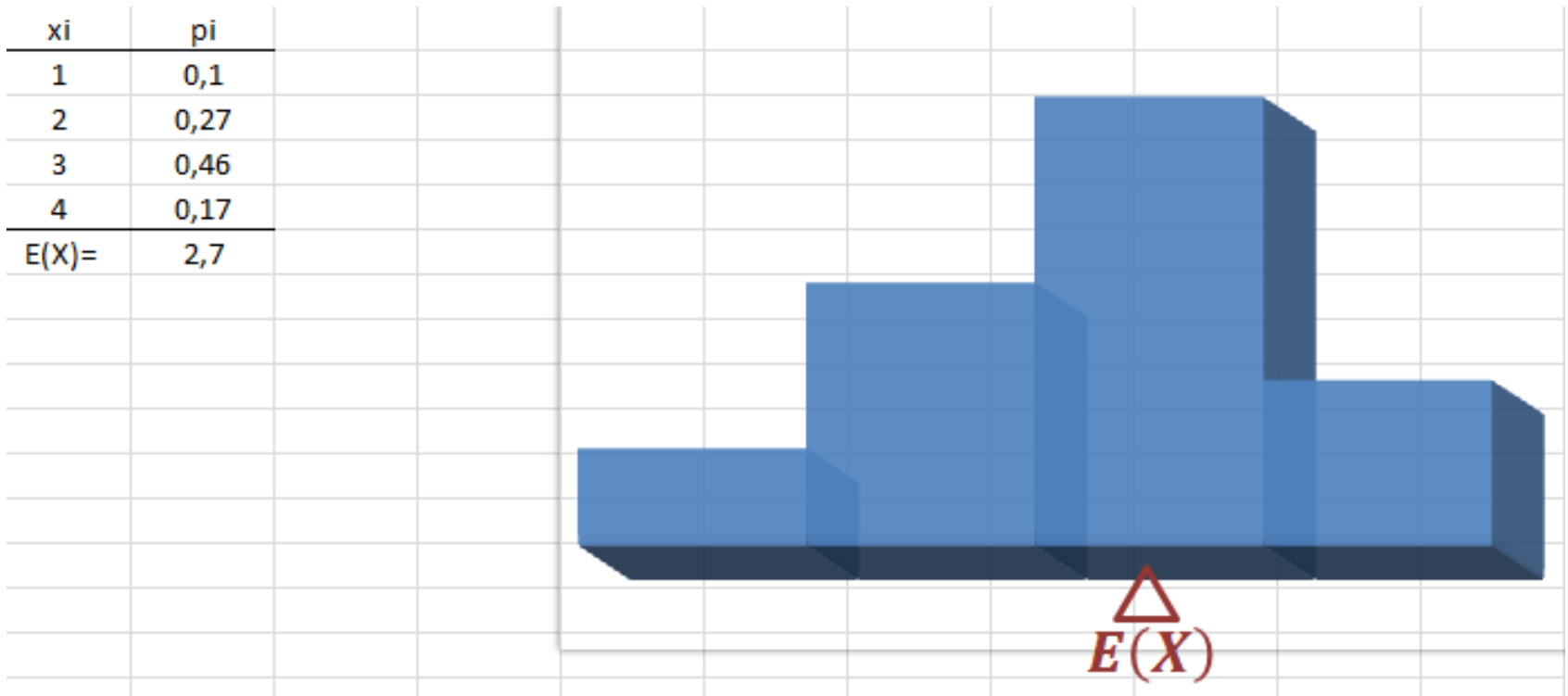
- pre diskrétnu náhodnú premennú: $E(X) = \sum x_i p_i$,
- pre spojitú náhodnú premennú $E(X) = \int_{-\infty}^{\infty} x f(x) dx$
(ak rad, resp. integrál existuje).

Význam pojmu:

$E(X)$ náhodnej premennej X je možné chápať ako:

- priemernú (očakávanú) hodnotu náhodnej premennej X , okolo ktorej hodnoty náhodnej premennej kolíšu,
- vážený priemer všetkých možných hodnôt ($E(X) = \sum_{(i)} x_i \cdot P(x_i)$),
- x -ovú hodnotu ťažiska hmotnej plochy pod krivkou hustoty rozdelenia pravdepodobnosti pre SNP - „ťažisko“ možných hodnôt.

Stredná hodnota $E(X)$ náhodnej premennej X ako „ťažisko“ možných hodnôt.



Vlastnosti strednej hodnoty:

$$E(c) = c$$

$$E(aX + b) = aE(X) + b$$

$$E(X + Y) = E(X) + E(Y)$$

Modus – mode – $Mo(X)$

- pre diskrétnu náhodnú premennú X je tá **hodnota náhodnej premennej**, ktorá má **najväčšiu pravdepodobnosť**, t.j. je to najpravdepodobnejšia hodnota X .
- pre spojitú náhodnú premennú je to **číslo**, v ktorom má jej funkcia hustoty pravdepodobnosti $f(x)$ **ostré lokálne maximum**.

Hovoríme, že náhodná premenná má rozdelenie

antimodálne rozdelenie – ak nemá modus;

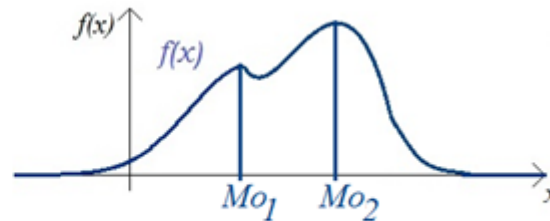
monomodálne rozdelenie – ak má jeden modus;

bimodálne rozdelenie – ak má dva modusy;

polymodálne rozdelenie – ak existujú viac ako dva modusy.



Antimodálne rozdelenie

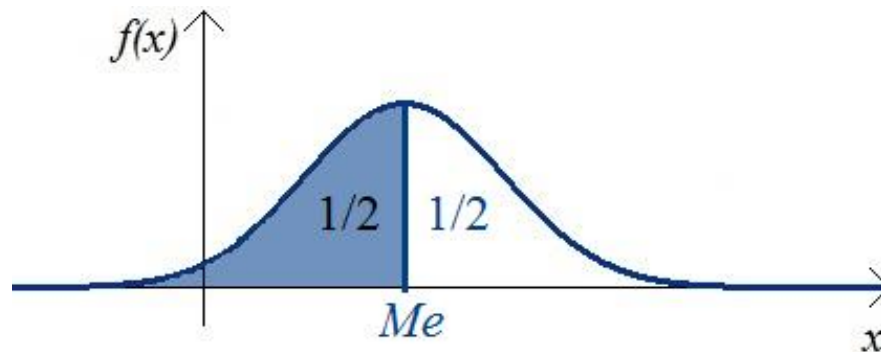


Bimodálne rozdelenie

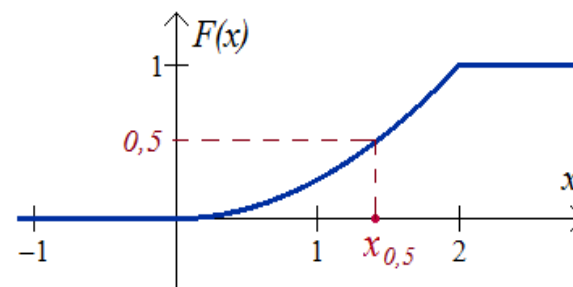
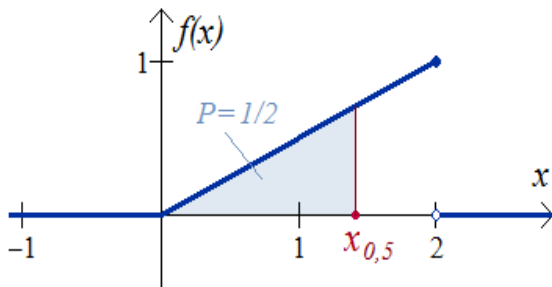
Medián $Me(X)$

spojitej náhodnej premennej X je hodnota, pre ktorú platí

$$F(Me) = P(X \leq Me) = \int_{-\infty}^{Me} f(x) dx = \frac{1}{2}$$



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

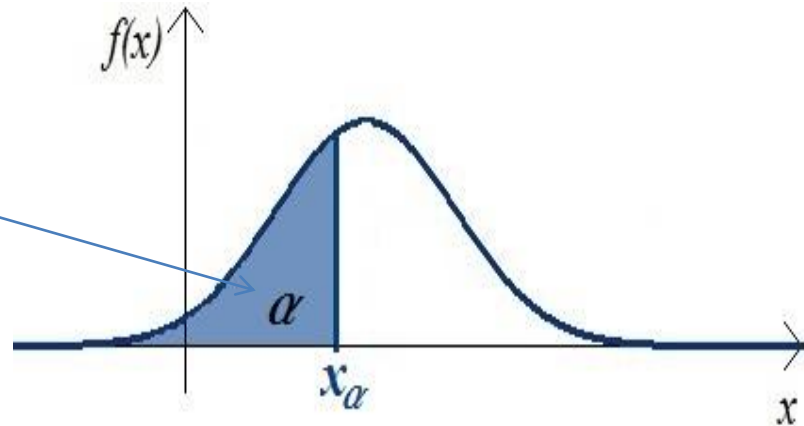


Kvantily spojitej náhodnej premennej X

Kvantilová funkcia $F^{-1}(x)$ spojitej náhodnej premennej X je inverzná funkcia k tej časti tejto distribučnej funkcie $F(x)$, ktorá je rastúca na množine M a zobrazuje ju na interval $(0,1)$.

α -kvantil spojitej náhodnej premennej X je číslo $x_\alpha = F^{-1}(\alpha)$, kde $\alpha \in (0,1)$.

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



Pre α -kvantil (používa sa tiež $\alpha \cdot 100\%$ kvantil) náhodnej premennej X platí

$$P(X \leq x_\alpha) = F(x_\alpha) = \alpha \Leftrightarrow x_\alpha = F^{-1}(\alpha)$$

Význačné kvantily náhodnej premennej X

kvartily

prvý alebo dolný kvartil $x_{0,25}$

medián $x_{0,5}$

tretí alebo horný kvartil $x_{0,75}$

decily

prvý decil $x_{0,1}$, druhý decil $x_{0,2}$, ..., deviaty decil $x_{0,9}$

percentily

prvý percentil $x_{0,01}$, druhý percentil $x_{0,02}$...,

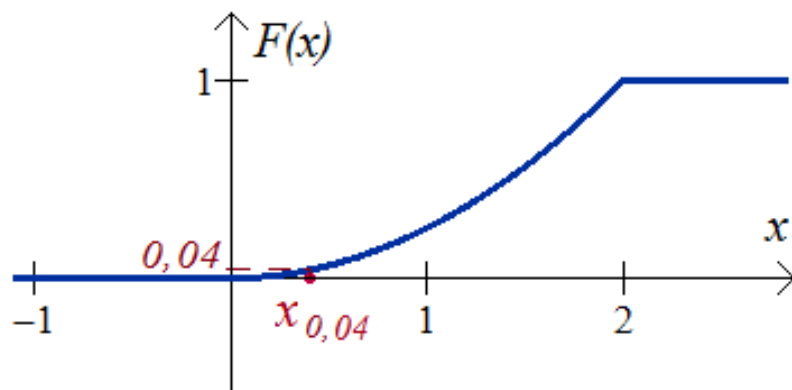
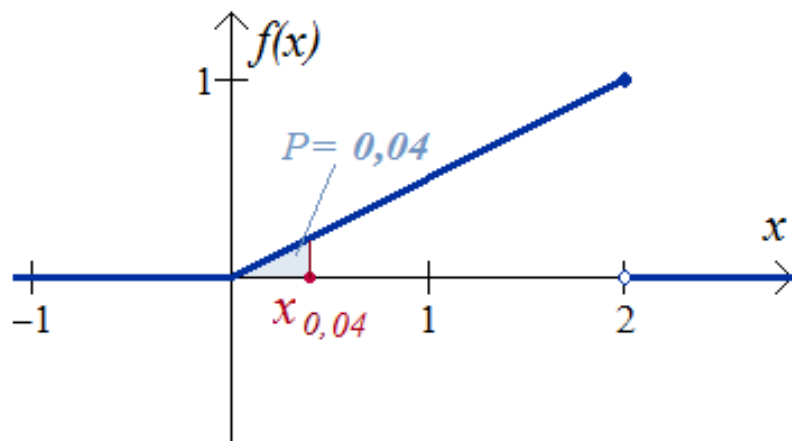
deväťdesiaty deviaty percentil $x_{0,99}$

Všeobecný názov kvantil spojitej NP zahŕňa rôzne pomenovania podľa počtu delení plochy pod krivkou hustoty na rovnako veľké časti:

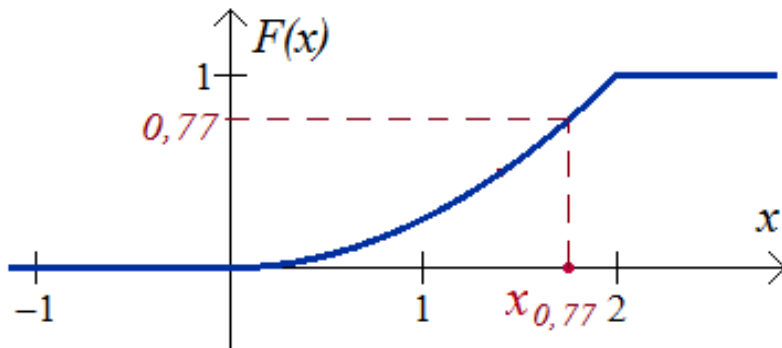
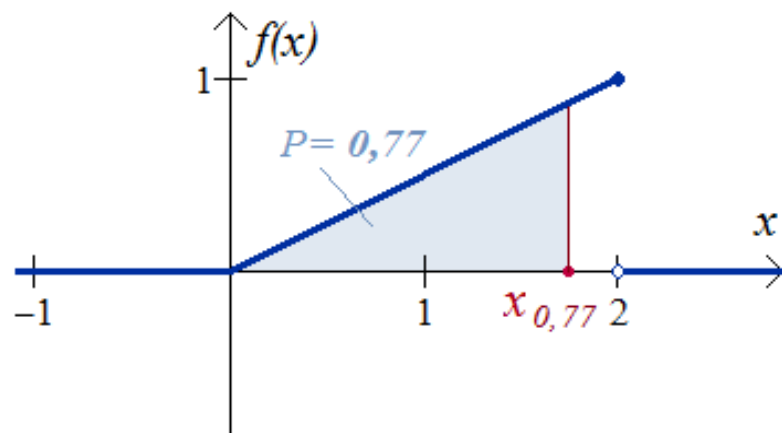
median (2quantile), tercily/tertily (3), quartily (4), quintily (5), sextily (6), septily (7), octily (8), decily (10), hexadecily (16), ventily/vigintily (20), centily/percentily (100)

Zobrazenie vybraných percentilov na grafe funkcie hustoty pravdepodobnosti $f(x)$ a na grafe distribučnej funkcie $F(x)$

4. percentil



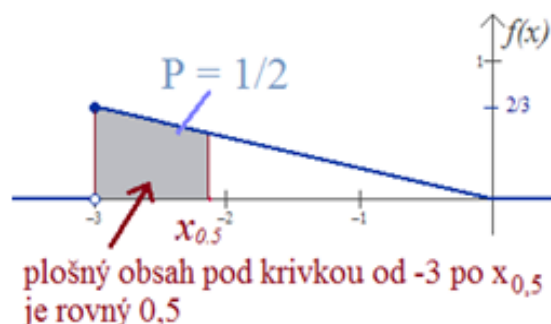
77. percentil



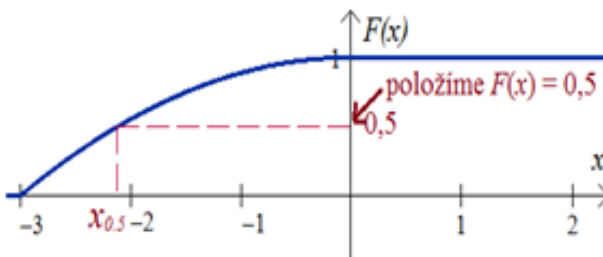
Zobrazenie mediánu na grafe funkcie hustoty pravdepodobnosti $f(x)$, na grafe distribučnej funkcie $F(x)$ a na grafe kvantilovej funkcie $F^{-1}(\alpha)$



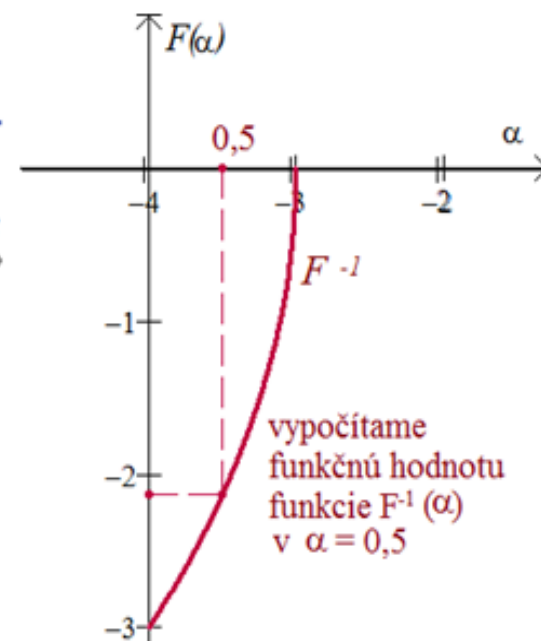
Me vyjadrený pomocou plošného obsahu pod krivkou hustoty pravdepodobnosti $f(x)$:



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

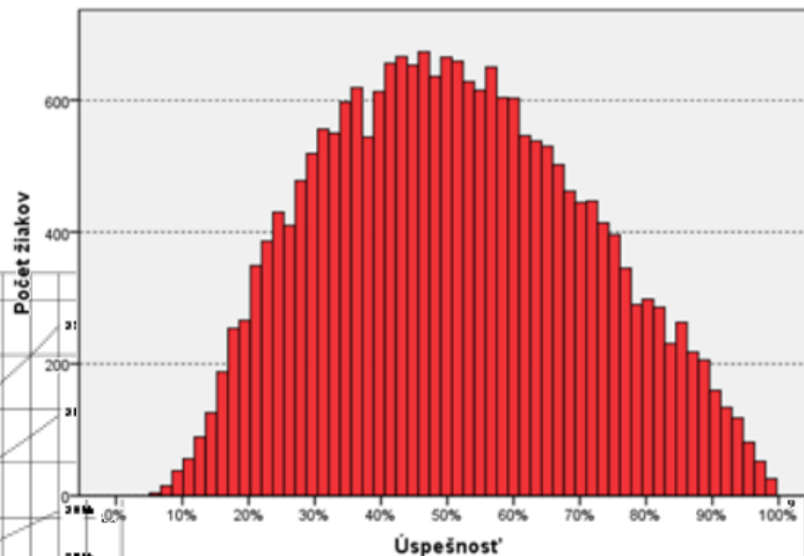


Me vyjadrený pomocou kvantilovej funkcie $F^{-1}(\alpha)$:

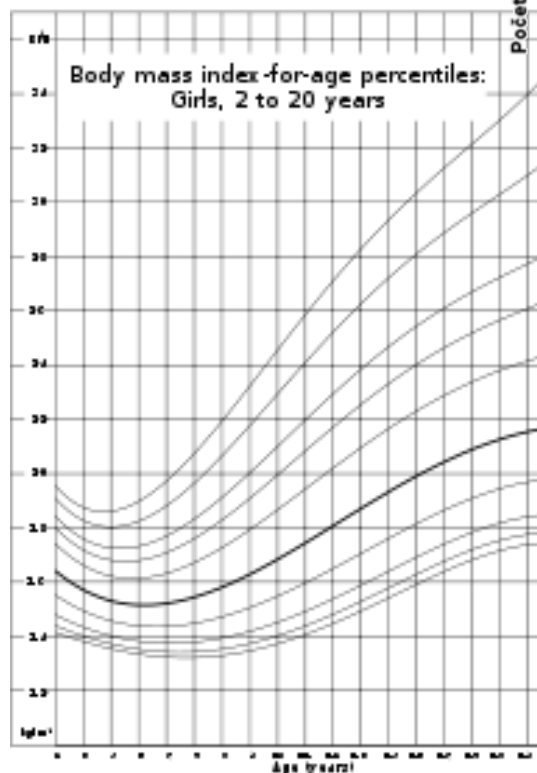


Kde všude sa môžeme
v praxi stretnúť s
kvantilmi?

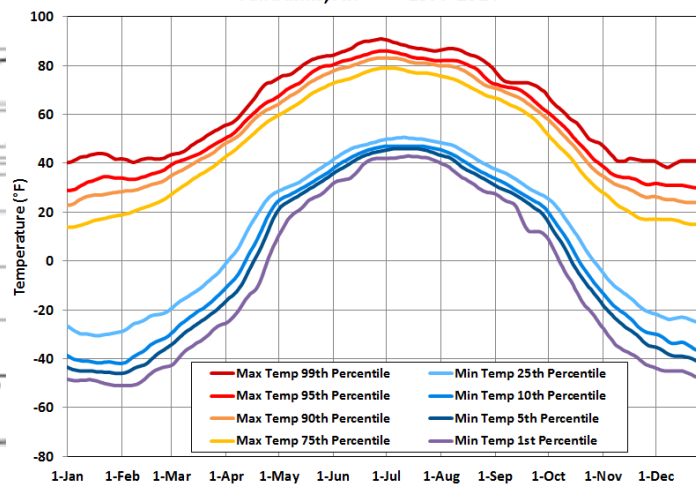
Anglický jazyk úroveň B1 2015 - histogram úspešnosti



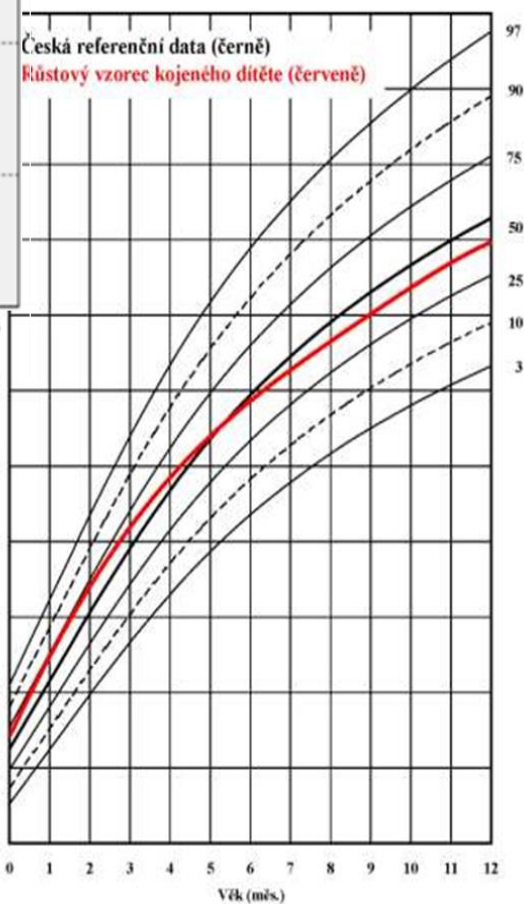
CDC Growth Charts United States



Daily Maximum and Minimum Temperature Percentiles
Fairbanks, AK 1977-2014



HMOTNOST - schematický graf



Charakteristiky variability náhodnej premennej X

Disperzia – rozptyl – VARIANCE – $D(X)$ určuje rozptýlenosť hodnôt náhodnej premennej X okolo jej strednej hodnoty s prihliadnutím na ich pravdepodobnosť.

Je to priemer druhých mocnín odchýlok od priemeru, t.j. stredná kvadratická odchýlka od strednej hodnoty: $D(X) = E \left[(X - E(X))^2 \right] = E(X^2) - E^2(X)$

- pre diskretnú náhodnú premennú $D(X) = \sum_i (x_i - E(X))^2 p_i$,
- pre spojitú náhodnú premennú $D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$.

Jednotka rozptylu je kvadrátom jednotky náhodnej premennej!!!

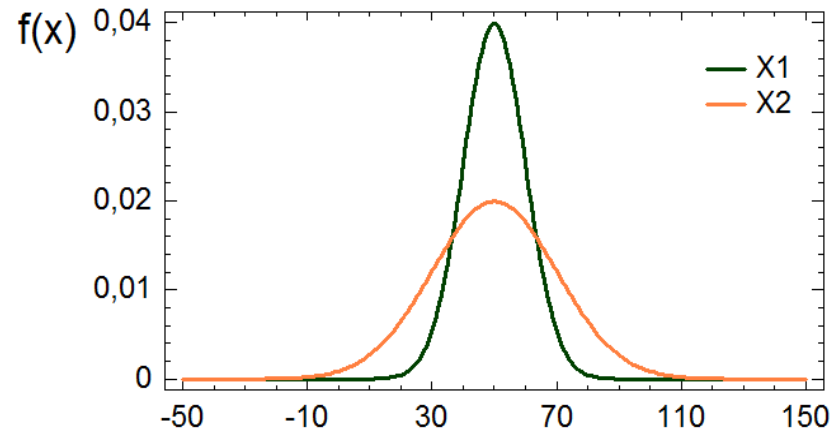
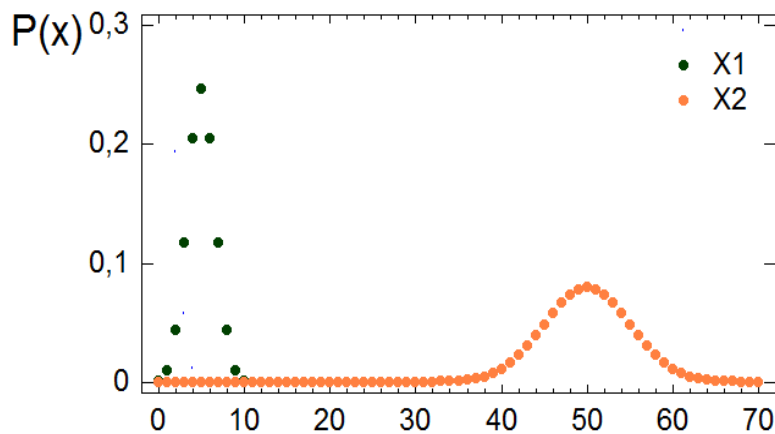
Vlastnosti disperzie:

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

Čím väčšia je disperzia, tým viac hodnoty náhodnej premennej X kolíšu okolo jej strednej hodnoty, t.j. stredná hodnota má horšiu vypovedaciu schopnosť o vlastnostiach náhodnej premennej X .

Malý rozptyl $D(X) \approx$ hodnoty náhodnej premennej X sa s vysokou pravdepodobnosťou nachádzajú blízko $E(X)$.

Veľký rozptyl $D(X) \approx$ hodnoty náhodnej premennej X sa často objavujú vo veľkej vzdialenosti od $E(X)$



$$D(X_1) < D(X_2)$$

Štandardná (smerodajná) odchýlka – STANDARD DEVIATION –

$\sigma(X)$ - druhá odmocnina z disperzie náhodnej premennej X

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

Smerodajná odchýlka

- je vyjadrená v rovnakých jednotkách ako náhodná premenná X ;
- neumožňuje porovnávať variabilitu náhodných premenných vyjadrených v rôznych jednotkách.

Príklad.

Hrám hru s jedným protihráčom. Ak na kocke padne 1, súper mi zaplatí 4 €, ak padne 2 alebo 3, súper mi zaplatí 1€, ak padne 4, súper mi zaplatí 2€, v ostatných prípadoch ja platím súperovi 3€.

Je pre mňa táto hra výhodná? Aký mám priemerný zisk (stratu) po jednom hode? Akú štandardnú odchýlku má môj zisk?

Riešenie:

X – výška mojej výhry

$M = \{-3; 1; 2; 4\}$

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

$$E(X) = \sum_i x_i p_i \Rightarrow E(X) = \frac{2}{6} \text{ €}$$

$$D(X) = \sum_i (x_i - E(X))^2 p_i \Rightarrow D(X) = 6,56 \text{ €}^2$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\sum_i (x_i - E(X))^2 p_i} \Rightarrow \sigma(X) = 2,56 \text{ €}$$

Príklad.

Nech $f(x) = \begin{cases} cx & \text{pre } x \in \langle -2, 0 \rangle; \\ 0 & \text{pre } x \notin \langle -2, 0 \rangle \end{cases}$ je hustota rozdelenia pravdepodobnosti

spojitej náhodnej premennej X . Nájdite číslo c , modus, disperziu a smerodajnú odchýlku náhodnej premennej X . Nájdite a znázornite distribučnú a kvantilovú funkciu náhodnej premennej X . Nájdite a graficky znázornite nasledujúce kvantily: medián, $x_{0,04}$, $x_{0,77}$.

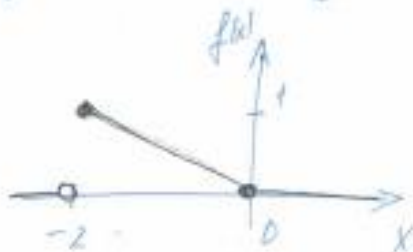
Riešenie:

$$f(x) = \begin{cases} cx, & x \in \langle -2, 0 \rangle \\ 0, & x \notin \langle -2, 0 \rangle \end{cases}$$



$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad \left\{ \begin{array}{l} \int_{-\infty}^{-2} 0 dx + \int_{-2}^0 cx dx + \int_0^{\infty} 0 dx = \left[c \cdot \frac{x^2}{2} \right]_{-2}^0 = -2c = 1 \Rightarrow c = -\frac{1}{2} \end{array} \right.$$

$$f(x) = \begin{cases} -\frac{x}{2}, & x \in \langle -2, 0 \rangle \\ 0, & x \notin \langle -2, 0 \rangle \end{cases}$$



$$S_0 = \frac{a \cdot \pi}{2} \Rightarrow S_0 = \frac{2 \cdot 1}{2}$$

$$M_0(X) = -2$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \Rightarrow E(X) = \int_{-\infty}^{-2} x \cdot 0 dx + \int_{-2}^0 x \cdot \left(-\frac{x}{2}\right) dx + \int_0^{\infty} x \cdot 0 dx = \int_{-2}^0 \left(-\frac{x}{2}\right) dx =$$

$$= \left[-\frac{x^2}{2} \right]_{-2}^0 = 0 - \left(-\frac{(-2)^2}{2} \right) = -\frac{4}{2} = -2$$

$$D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx = \dots \int_{-2}^0 \left(x + \frac{4}{2}\right)^2 \cdot \left(-\frac{x}{2}\right) dx = \int_{-2}^0 \left(x^2 + 4x + \frac{16}{2}\right) \cdot \left(-\frac{x}{2}\right) dx =$$

$$= -\frac{1}{2} \int_{-2}^0 \left(x^3 + 4x^2 + 8x\right) dx = -\frac{1}{2} \left[\frac{x^4}{4} + \frac{4x^3}{3} + \frac{8x^2}{2} \right]_{-2}^0 = -\frac{1}{2} \left(0 - \left(\frac{(-2)^4}{4} + \frac{4 \cdot (-2)^3}{3} + \frac{8 \cdot (-2)^2}{2} \right) \right)$$

$$= \frac{1}{2} \left(\frac{16}{4} - \frac{64}{3} + \frac{32}{1} \right) = \frac{1}{2} \cdot \frac{16}{3} = \frac{8}{3} = \frac{2}{3}$$

$$D(X) = E(X^2) - E^2(X)$$

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

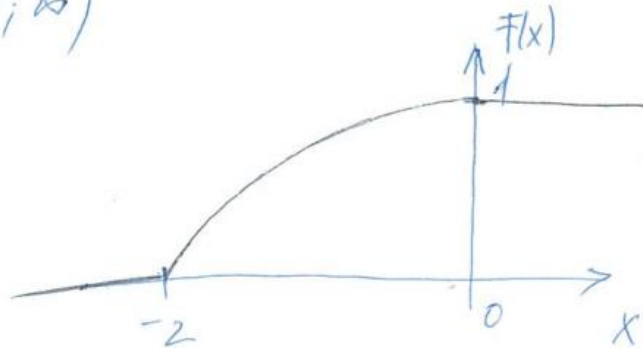
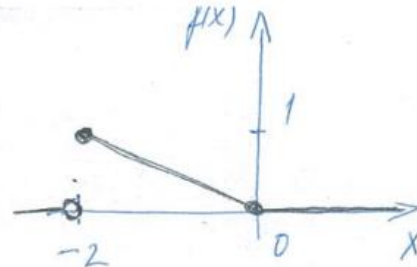
$$D(X) = 2 - \left(-\frac{4}{2}\right)^2 = 2 - \frac{16}{4} = \frac{2}{4} = \frac{1}{2}$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{1}{2}}$$

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

$$f(x) = \begin{cases} 0 & ; x \in (-\infty; -2) \end{cases}$$

$$\begin{cases} \int_{-2}^x -\frac{1}{2} dt = -\frac{1}{2} \left[\frac{t^2}{2} \right]_{-2}^x = -\frac{1}{4} \left[t^2 \right]_{-2}^x = -\frac{1}{4} (x^2 - (-2)^2) = -\frac{1}{4} x^2 + 1 & ; x \in [-2; 0] \\ 1 & ; x \in [0; \infty) \end{cases}$$



$$y = ax^2 + bx + c$$

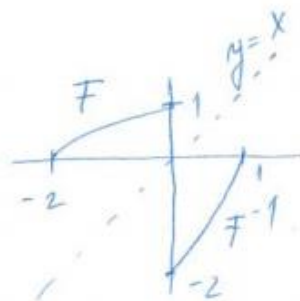
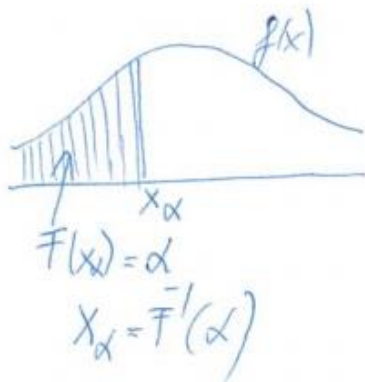
$$a > 0$$



$$a < 0$$



$$F^{-1}(\alpha) = ? \quad \alpha \in (0, 1)$$



$$\alpha = -\frac{1}{4}x_\alpha^2 + 1$$

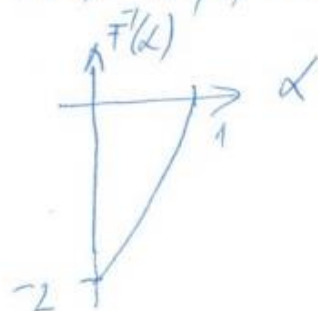
$$4\alpha = -x_\alpha^2 + 4$$

$$x_\alpha^2 = 4 - 4\alpha$$

$$x_\alpha = \pm \sqrt{4 - 4\alpha} \Rightarrow F^{-1}(\alpha) = -\sqrt{4 - 4\alpha}, \quad \alpha \in (0, 1)$$

$$\mathcal{D}(F) = (-2, 0) = \mathcal{H}(F^{-1})$$

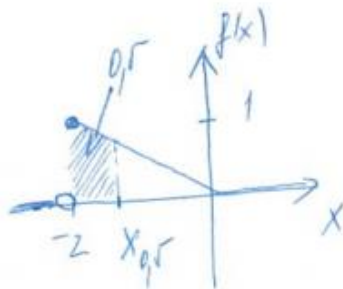
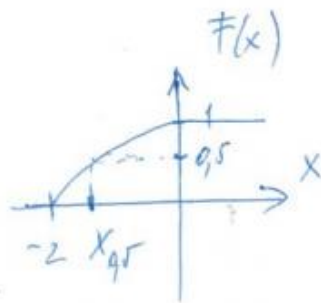
$$\mathcal{H}(F) = (0, 1) = \mathcal{D}(F^{-1})$$



$$x_{0,5} = -\sqrt{4 - 4 \cdot \frac{1}{2}} = -\sqrt{2}$$

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

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



Všeobecné charakteristiky – momenty

Ak $k \in N$, p_i je pravdepodobnostná funkcia diskkrétnej náhodnej premennej, $f(x)$ je hustota pravdepodobnosti spojitej náhodnej premennej, potom

k -ty začiatočný moment náhodnej premennej X

- pre diskrétnu náhodnú premennú X : $z_k(X) = \sum_i x_i^k \cdot p_i$,
- pre spojitú náhodnú premennú X : $z_k(X) = \int_{-\infty}^{\infty} x^k f(x) dx$.

k -ty centrálny moment náhodnej premennej X

- pre diskrétnu náhodnú premennú X : $c_k(X) = \sum_i (x_i - E(X))^k \cdot p_i$,
- pre spojitú náhodnú premennú X : $c_k(X) = \int_{-\infty}^{\infty} (x - E(X))^k f(x) dx$.

Využitie všeobecných charakteristík - momentov

Stredná hodnota $E(X)$ je **prvý** začiatočný moment

- pre diskretnú náhodnú premennú X : $E(X) = z_1(X) = \sum_i x_i \cdot p_i$,
- pre spojitú náhodnú premennú X : $E(X) = z_1(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$.

Rozptyl $D(X)$ je **druhý** centrálny moment

- pre diskretnú náhodnú premennú X : $D(X) = c_2(X) = \sum_i (x_i - E(X))^2 \cdot p_i$,
- pre spojitú náhodnú premennú X : $D(X) = c_2(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$.

Koeficient šikmosti a špicatosti

Vzťahy medzi centrálnymi a začiatočnými momentami:

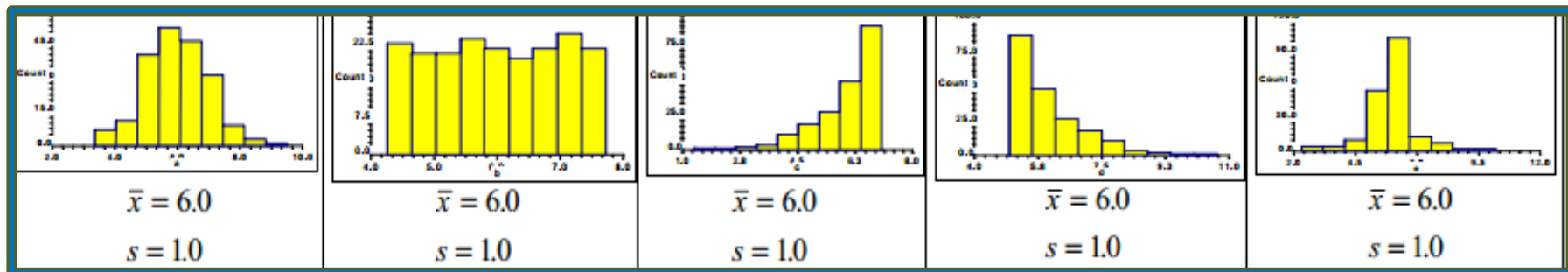
$$c_1 = 0;$$

$$c_2 = z_2 - z_1^2;$$

$$c_3 = z_3 - 3z_2z_1 + 2z_1^3;$$

$$c_4 = z_4 - 4z_3z_1 + 6z_2z_1^2 - 3z_1^4.$$

Sú miery polohy a variability dostatočné pre posúdenie rozdelenia sledovaných premenných?

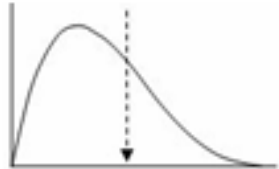
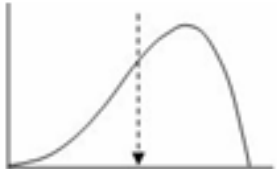


Všetkých 5 ukážok má rovnaké charakteristiky polohy i variability (priemery i smerodajné odchýlky sú zhodné), ale vidíme, že tvary rozdelení sú rôzne.

Charakteristiky symetrie náhodnej premennej X

Koeficient šikmosti – skewness – $\alpha(X)$ charakterizuje symetriu rozdelenia náhodnej premennej X :

$$\alpha(X) = \frac{c_3}{\sigma^3(X)} = \frac{c_3}{\sqrt{D^3(X)}}$$

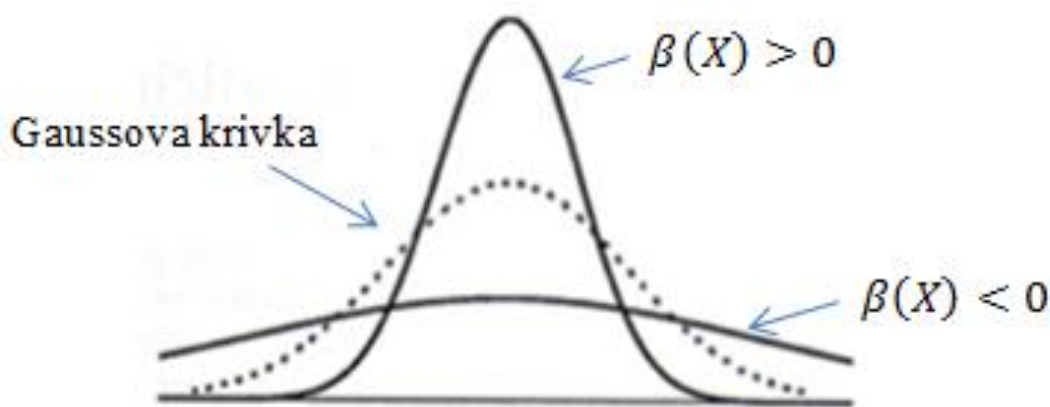
Graf funkcie hustoty je			
	symetrický	nesymetrický	
		zošikmenie doprava (vrchol vpredu) 	zošikmenie doľava (vrchol vzadu) 
koeficient šikmosti $\alpha(X)$ je	0	> 0 (kladný)	< 0 (záporný)
vzťah medzi $E(X)$, Mo , Me	$E(X) = Mo = Me$	$Mo < Me < E(X)$ (častejší výskyt menších hodnôt)	$E(X) < Me < Mo$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

Charakteristiky koncentrácie náhodnej premennej X

Koeficient špicatosti – kurtosis – $\beta(X)$ špicatosť, strmosť rozdelenia náhodnej premennej X :

$$\beta(X) = \frac{c_4}{\sigma^4(X)} - 3 = \frac{c_4}{D^2(X)} - 3$$

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



Rozhodnite, ktoré z tvrdení sú pravdivé:

1. Ak je rozdelenie náhodnej premennej symetrické, potom $\alpha(X) = 0$.
2. Ak je rozdelenie náhodnej premennej symetrické, potom $E(X) = x_{0,5}$.
3. Ak má náhodná premenná symetrické rozdelenie s jedným modulusom, potom je jej stredná hodnota rovná mediánu a zároveň modusu.
4. Stredná hodnota konštanty je rovná tejto konštante.
5. Rozptyl konštanty je rovný nule.