Katedra štatistiky,
Fakulta hospodárskej informatiky,
Ekonomická univerzita v Bratislave

Pravdepodobnostné modelovanie inverznými distribučnými funkciami: Charakteristiky kvantilových rozdelení

Ľubica SIPKOVÁ

December 2009

5. z cyklu prezentácií

Rôzna báza konštrukcie štatistík

Vlastnosti rozdelenia možno charakterizovať štatistikami na rôznej báze:

- momentov
- kvantilov
- lineárnych momentov
- pravdepodobnostných momentov

Stredná hodnota vyjadrená pomocou QF

- definovaného pomocou funkcie hustoty $f(x)$

(Báza-momentová):

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

- Dosadením $x = Q(p)$ a d$p = f(x)$ dx pre $0 \le p \le 1$

(Báza-kvantilová):

$$E(X) = \int_{0}^{1} Q(p)\,\mathrm{d}p \qquad \text{pre} \quad 0 \le p \le 1$$

pre rozdelenia s predĺženým koncom je stredná hodnota definovaná len pre niektoré hodnoty parametrov QF

Stredná hodnota Paretovho rozdelenia

$S(p) = 1/(1-p)^{\beta}$, kde $\beta > 0,\ 0 \le p < 1$

- napr. Paretovo rozdelenie s dlhou pravou stranou má konečnú strednú hodnotu len pre $0 < \beta < 1$

$$E(X) = \int_0^1 Q(p)\,\mathrm{d}p$$

- pre $\beta > 1$ platí $E(X) = \infty$

k-ty začiatočný moment pomocou kvantilovej funkcie

- pomocou funkcie hustoty $f(x)$ (Báza-momentová):

$$m_k = \int_{-\infty}^{\infty} x^k f(x)\,\mathrm{d}x$$

k-ty začiatočný moment $m_k = E(X^k)$

- **pomocou kvantilovej funkcie** (Báza-kvantilová):

$$\boldsymbol{m_k} = \int_0^1 [Q(p)]^k \,\mathrm{d}p$$

k-ty centrálny moment pomocou kvantilovej funkcie

- pomocou funkcie hustoty *f*(*x*) (Báza-momentová):

$$\mu_k = \int_{-\infty}^{\infty} \left[x - E(X)\right]^k f(x)\,\mathrm{d}x$$

k-ty centrálny moment

- pomocou kvantilovej funkcie (Báza-kvantilová) :

$$\mu_k = \int_{0}^{1} \left[Q(p) - E(X)\right]^k \mathrm{d}p$$

Dosadením za *k* možno získať tvar všetkých základných momentových charakteristík priamo z kvantilovej funkcie

Rozptyl pomocou kvantilovej funkcie

- *k*-ty centrálny moment

$$\mu_k = \int_0^1 \left[Q(p) - E(X)\right]^k \,\mathrm{d}p$$

- **druhý centrálny moment**

$$D(X) = \mu_2 = E\left[\left(X - E(X)\right)^2\right] = \int_0^1 \left[Q(p) - E(X)\right]^2 \,\mathrm{d}p$$

Stredná hodnota transformovanej funkcie

$$E[T(X)] \neq T[E(X)]$$

- t. j. stredná hodnota transformovanej funkcie nie je transformácia integrálu, ale integrál transformácie (spôsobuje to problémy pri stanovení strednej hodnoty po úpravách QF)

$$E\left[T\left(X\right)\right] = \int_{0}^{1} T\left[Q\left(p\right)\right] \mathrm{d}p$$

Kvantilové charakteristiky QF

- je vhodnejšie charakterizovať vlastnosti kvantilového rozdelenia pomocou **kvantilových charakteristík**

  (kvantilové charakteristiky kvantilových modelov)

- sú analógiou výberových kvantilových charakteristík
- **vo vzťahoch ich definovania**

namiesto výberovej kvantilovej funkcie $\tilde{Q}(p)$

je použitá kvantilová funkcia rozdelenia $Q(p)$

Nasledujú definície kvantilových charaktristík s ich schématickým zobrazením

Výberové kvantilové charakteristiky polohy:

- **Označenie:**

minimálna hodnota (s), dolný kvartil (lq), medián (m), horný kvartil (uq), maximálna hodnota (l)

sú to výberové p-kvantily pre p = 0; 0,25; 0,5; 0,75; 1

a sú východiskom najznámejších výberových charakteristík variability a šikmosti znázornené v Box-Plot-e

- **Pre ľubovolné p :** **výberové p - kvantily**

percentily, decily, oktily, sixtily

Výpočet p – kvantilov

p-kvantily možno počítať cez funkciu excelu :

PERCENTILE (array;p)

Podľa Parzena je empirická kvantilová distribučná funkcia, definovaná:

$$\tilde{Q}(p_r) = x_r\,, \quad \text{pre } (r-1)/n < p_r \leq r/n$$

je inverznou funkciou k $\tilde{F}(x_r)$, neklesajúcou a spojitou len zľava v každom $p_r = (r - 0{,}5)/n$

rovnajúca sa hodnote kvantilu X_r pre $r = 1, 2, \ldots, n$

Empirickú kvantilovú distribučnú funkciu $\tilde{Q}(p)$ pre spojitú náhodnú premennú možno pomocou interpolácie zapísať:

$$\tilde{Q}(p) = (1 - z)\, x_s + z\, x_{s+1}\,,$$

pre $(r - 1)/n < p \leq r/n$, kde s je celočíselná časť r,

$r = np + 0{,}5 \qquad z = r - s$

Výpočet *p*-kvantilov v EXCEL-i

Microsoft Excel - kv. charakteristiky 4.cvicenie KMaS

Súbor Úpravy Zobraziť Vložiť Formát Nástroje Údaje Okno Pomocník

100% Arial 8

G2 =VLOOKUP(F2;$A$2:$F$76;2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	r	x	pr=(r-0,5)/n	r=n*pr+0,5	p	s	x(s)	x(s+1)	z=r-s	1-z	empQ(p)	abs med d		med med d
2	1	19928,00	0,0067	1,2500	0,010	1	19928,00	20412,00	0,2500	0,7500	20049,0000	40108,1400	ABS(B2-$K$8)	0,0000
3	2	20412,00	0,0200	4,2500	0,050	4	21976,00	22387,00	0,2500	0,7500	22078,7500	39624,1400		317,6000
4	3	21571,25	0,0333	8,0000	0,100	8	32702,00	42252,00	0,0000	1,0000	32702,0000	38464,8864		728,0000
5	4	21976,00	0,0467	9,8750	0,125	9	42252,00	42432,06	0,8750	0,1250	42409,5516	38060,1400		1342,0000
6	5	22387,00	0,0600	19,2500	0,250	19	54712,00	55009,55	0,2500	0,7500	54786,3867	37649,1400		1410,0000
7	6	31350,70	0,0733	28,6250	0,375	28	58956,00	59700,00	0,6250	0,3750	59421,0000	28685,4400		1496,0000
8	7	31620,00	0,0867	30,5000	0,400	30	60035,28	60037,00	0,5000	0,5000	60036,1400	28416,1400		1648,4133
9	8	32702,00	0,1000	34,2500	0,450	34	62504,00	62590,00	0,2500	0,7500	62525,5000	27334,1400		3496,0000
10	9	42252,00	0,1133	37,2500	0,490	37	63272,00	64000,00	0,2500	0,7500	63454,0000	17784,1400		3709,7200
11	10	42432,06	0,1267	38,0000	0,500	38	64000,00	64317,60	0,0000	1,0000	64000,0000	17604,0810		3963,0000
12	11	43686,00	0,1400	38,7500	0,510	38	64000,00	64317,60	0,7500	0,2500	64238,2000	16350,1400		3964,7200
13	12	45674,00	0,1533	41,7500	0,550	41	68270,00	68627,33	0,7500	0,2500	68538,0000	14362,1400		4270,0000
14	13	46134,00	0,1667	45,5000	0,600	45	69911,56	70028,55	0,5000	0,5000	69970,0534	13902,1400		4300,0000
15	14	46733,33	0,1800	47,3750	0,625	47	71041,34	71627,48	0,3750	0,6250	71261,1400	13302,8067		4627,3333
16	15	46973,00	0,1933	56,7500	0,750	56	83202,67	85483,87	0,7500	0,2500	84913,5699	13063,1400		4961,6720
17	16	53545,36	0,2067	66,1250	0,875	66	101776,42	129240,54	0,1250	0,8750	105209,4350	6490,7800		5044,0000
18	17	54001,60	0,2200	68,0000	0,900	68	130588,50	132279,92	0,0000	1,0000	130588,5017	6034,5400		5300,0000
19	18	54152,72	0,2333	71,7500	0,950	71	158385,48	177545,06	0,7500	0,2500	172755,1652	5883,4200		5397,8000
20	19	54712,00	0,2467	74,7500	0,990	74	230940,00	247991,81	0,7500	0,2500	243728,8612	5324,1400		5911,5600
21	20	55009,55	0,2600									5026,5933		6028,5467
22	21	55030,00	0,2733									5006,1400		6286,8000
23	22	55155,98	0,2867									4880,1600		7041,3360
24	23	55198,00	0,3000									4838,1400		7590,0000
25	24	55310,21	0,3133									4725,9267		7627,4800
26	25	56410,00	0,3267									3626,1400		8535,0000
27	26	57713,20	0,3400									2322,9400		8689,7867
28	27	58700,00	0,3533									1336,1400		8802,0000
29	28	58956,00	0,3667									1080,1400		8807,5040
30	29	59700,00	0,3800									336,1400		8844,0200

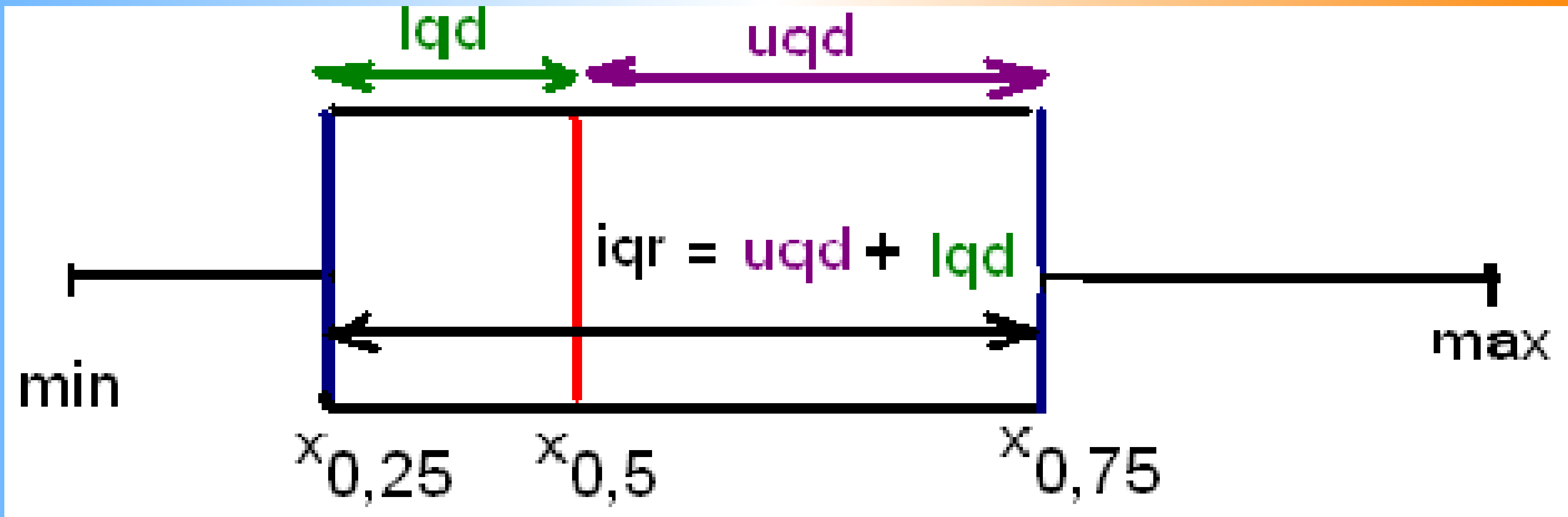
=VLOOKUP(F2;$A$2:$F$76;2)

=VLOOKUP(F2+1;$A$2:$F$76;2)

empQ(p) = (1-z)*x(s)+z*x(s+1)

Výberové charakteristiky variability (1):

- **horná kvartilová diferencia** $uqd = uq - m$
- **dolná kvartilová diferencia** $lqd = m - lq$
- **kvartilové rozpätie** $iqr = uqd + lqd = \tilde{Q}(1 - 0{,}25) - \tilde{Q}(0{,}25)$
- **kvartilová odchýlka** $qo(0{,}25) = iqr/2$



Výberové charakteristiky variability (2):

- **dolná *p*-diferencia** $ld(p) = m - \tilde{Q}(p)$,

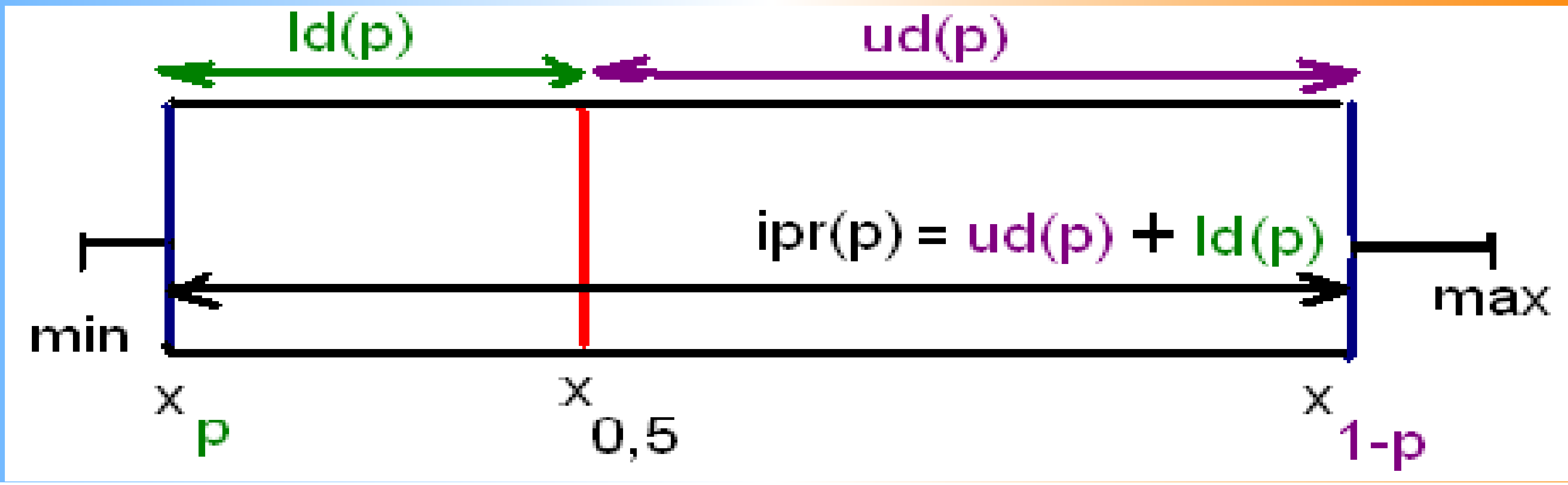
  kde $\tilde{Q}(p)$ je p-ty výberový kvantil a $0 \le p \le 0{,}5$

- **horná *p*-diferencia** $ud(p) = \tilde{Q}(1 - p) - m$,

  kde $\tilde{Q}(1 - p)$ je $(1 - p)$ -ty kvantil a $0 \le p \le 0{,}5$

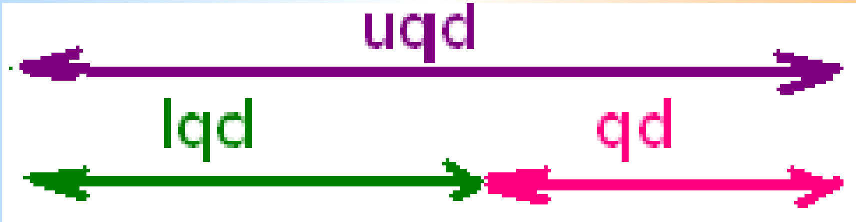
- **kvantilové rozpätie** $ipr(p) = ud(p) + ld(p) = \tilde{Q}(1 - p) - \tilde{Q}(p)$

- **kvantilová odchýlka** $qo(p) = ipr(p)/2$

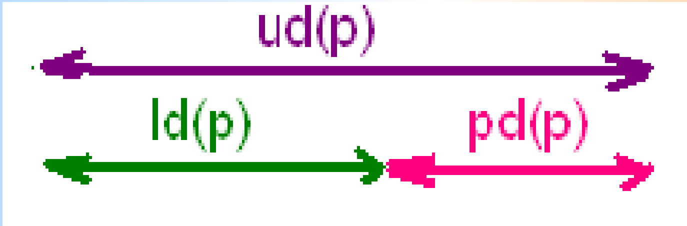


Výberové charakteristiky variability (3):

- **kvartilová diferencia** $qd = uqd - lqd$



- ***p*-diferencia** $pd(p) = ud(p) - ld(p)$



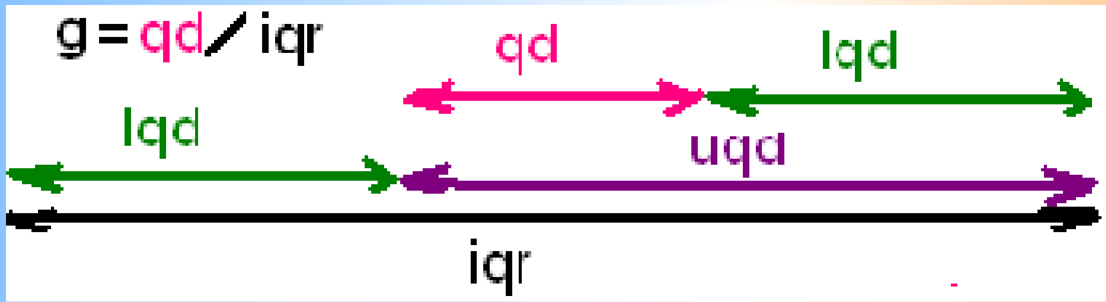
- **mediánová absolútna odchýlka**

$$\tilde{d} = \text{med}\left| x_i - \bar{x} \right|$$

- alebo $\tilde{d} = \text{med}\left| x_i - m \right|$

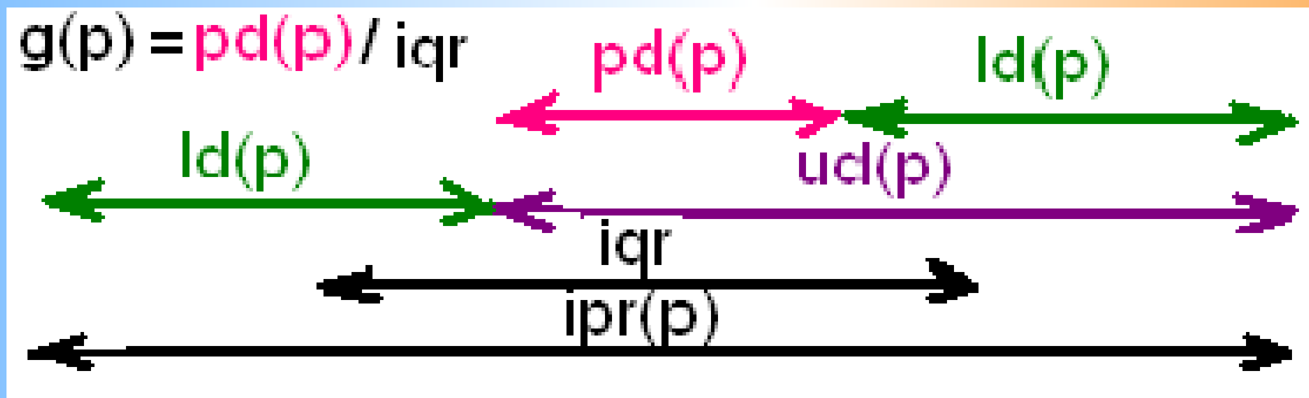
Výberové charakteristiky šikmosti (1):

- **kvartilová diferencia** $qd = lq + uq - 2m$
- **Galtonov koeficient šikmosti** $g = qd/iqr$



- **kvantilová diferencia, *p*-diferencia** $pd(p) = ud(p) - ld(p)$
- **Galtonova šikmosť** $g(p) = pd(p)/iqr$

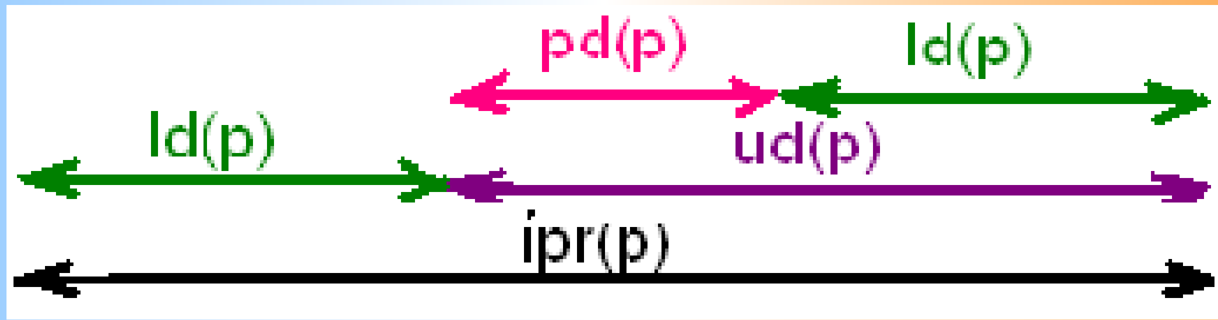
t. j. normovaná kvantilová diferencia pomocou kvartilového rozpätia



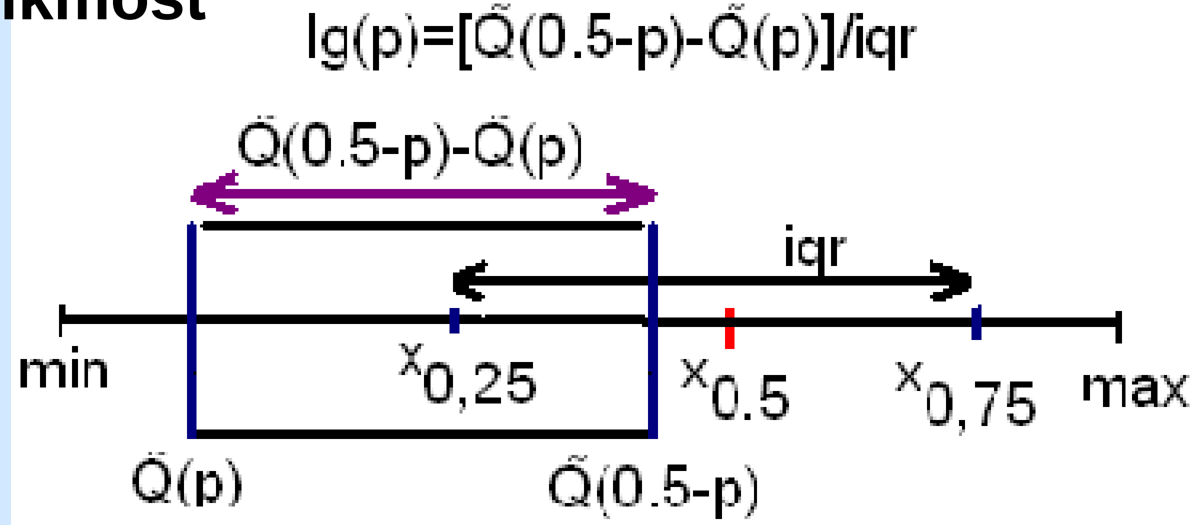
Výberové charakteristiky šikmosti:

- **Galtonov *p*-index šikmosti** $g^*(p) = pd(p)/ipr(p)$

t. j. štandardizovaná kvantilová diferencia pomocou kvantilového rozpätia



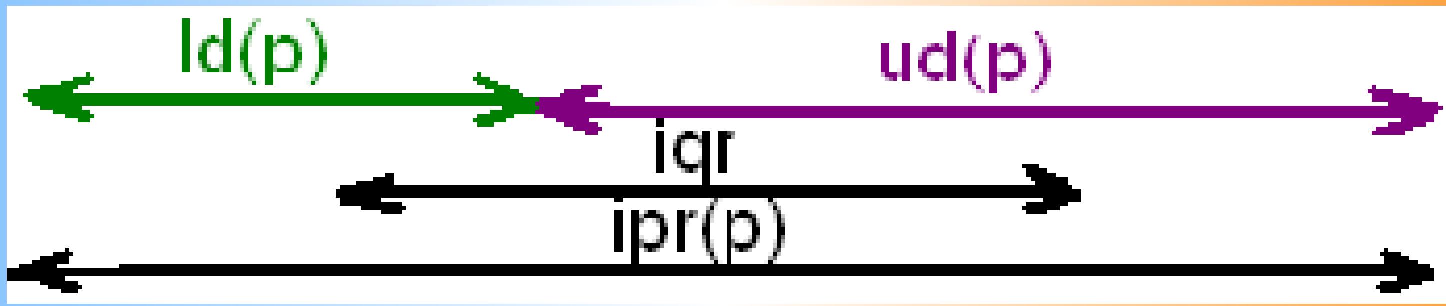
- **ľavá Galtonova *p*-šikmosť**



- **kvantilový podiel, kvantilová proporcia** $sr(p) = ud(p)/ld(p)$

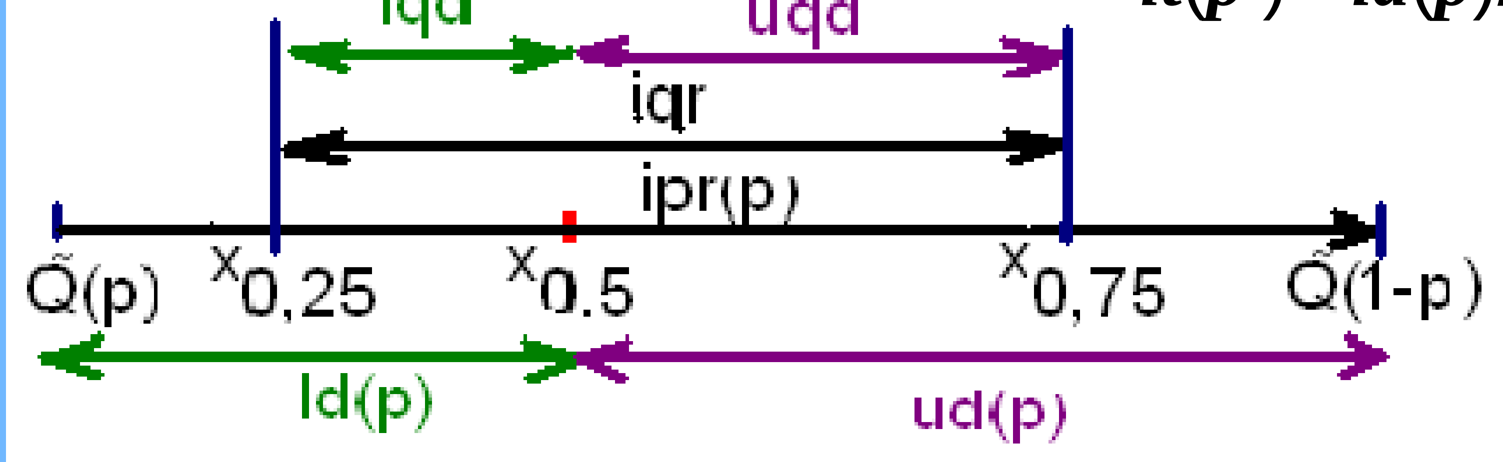
Výberové charakteristiky špicatosti:

- **index špicatosti** $t(p) = ipr(p)/iqr$, pre $0 \le p \le 0{,}5$



- **horný a dolný index špicatosti** $ut(p) = ud(p)/uqd$

$lt(p) = ld(p)/lqd$



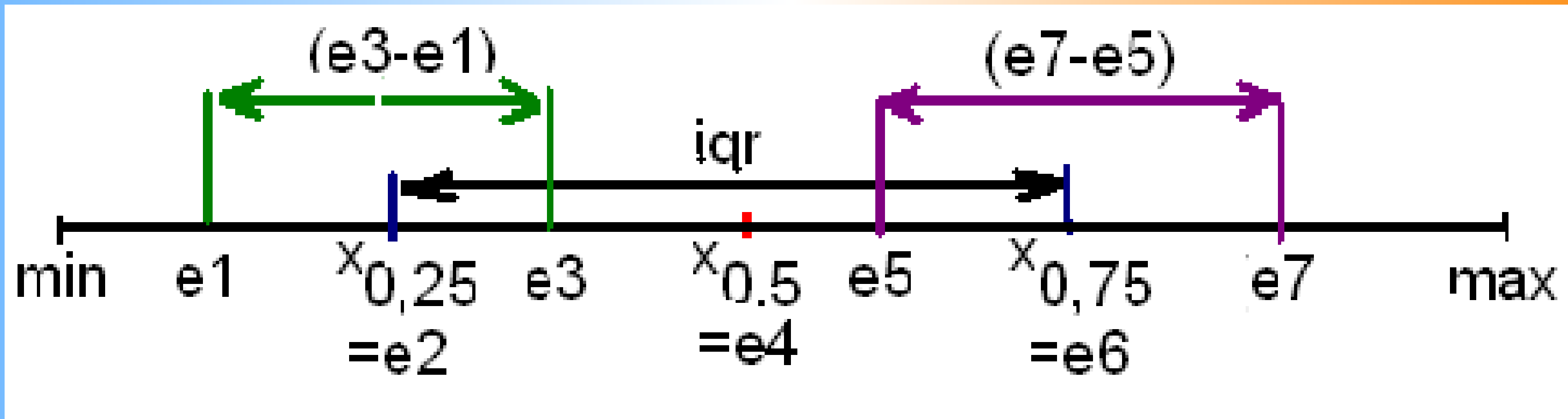
Výberové charakteristiky špicatosti:

- **Moorsova špicatosť k**

  vychádza z oktilov $e1$, $e2$,…, $e7$

  t. j. kvantilov deliacich súbor na osem častí

$$k = [(e7 - e5) + (e3 - e1)]/iqr$$



- **horná a dolná špicatosť**

$$uk(t) = [\tilde{Q}(1 - t) + \tilde{Q}(0,5 + t) - 2\tilde{Q}(0,75)]/[\tilde{Q}(1 - t) - \tilde{Q}(t)]$$

kde $0 < t < 0,25$

$$lk(t) = [\tilde{Q}(0,5 - t) + \tilde{Q}(t) - 2\tilde{Q}(0,25)]/[\tilde{Q}(1 - t) - \tilde{Q}(t)]$$

kde $0 < t < 0,25$

Moorsova sumarizácia o tvare rozdelenia

Štyri miery (m, iqr, g, k) :

1. **medián**
2. **kvartilové rozpätie**
3. **Galtonov koeficient šikmosti**
4. **Moorsova špicatosť**

poskytujú podľa **Moorsa** jednoduchú sumarizáciu o tvare rozdelenia na kvantilovom základe

Zodpovedá **Pearsonovej** štvorčíselnej sumarizácii – priemer, rozptyl, koeficient šikmosti a Pearsonova špicatosť na báze momentov.

Označenie kvantilových charakteristík QF

je adekvátne výberovým kvantilovým charakteristikám, no pre kvantilové charakteristiky teoretického rozdelenia, alebo rozdelenia základného súboru sa používajú **veľké písmená**

- Napríklad pre výberové kvantilové rozpätie $ipr(p)$:

$$ipr(p) = ud(p) + ld(p) = \tilde{Q}(1-p) - \tilde{Q}(p)$$

Kvartilové rozpätie teoretického rozdelenia, označené ako ***IPR(p)***, je definované ako

$$IPR(p) = UD(p) + LD(p) = Q(1-p) - Q(p)$$

Mediánové transformačné pravidlo

$$M[T(X)] = T[M(X)]$$

Medián (ľubovoľný kvantil) neklesajúcej transformačnej funkcie premennej *X* sa rovná adekvátnej transformácii mediánu (ľubovoľného kvantilu) tejto premennej.

Veľký rozsah uplatnenia pori rôznych transformáciách kvantilových funkcií

Pravdepodobnostné modelovanie inverznými distribučnými funkciami: Charakteristiky kvantilových rozdelení

Spracovanie piatej z cyklu prezentácií o kvantilovom modelovaní.

Podrobnejšie možno nájsť v monografii:

Sipková, Ľ; Sodomová, E.: Modelovanie kvantilovými funkciami, Vydavateľstvo EKONÓM, Bratislava, 2007; 175 s. (kap. 3.4; s.84 -91)

ISBN 978-80-225-2346-2

Ľubica SIPKOVÁ

december 2009