Leíró és matematikai statisztika

Matematika alapszak, matematikai elemző szakirány

Zempléni András

Valószínűségelméleti és Statisztika Tanszék
Matematikai Intézet
Természettudományi Kar
Eötvös Loránd Tudományegyetem

Honlap: `zempleni.elte.hu`

E-mail: `andras.zempleni@ttk.elte.hu`

Szoba: D 3-310

2. előadás

Érték-, ár- és volumenindexek I

Index vagy **indexszám**: közvetlenül nem összesíthető, de gazdaságilag összetartozó adatok átlagos változását mutató összetett viszonyszám.

Tegyük fel, hogy m különböző terméket értékesítünk két különböző időszakban, és az értékesítés árbevételét szeretnénk elemezni.

Jelölések:

- $q_{0,j}$: a j. termékből eladott mennyiség a bázisidőszakban
- $q_{1,j}$: a j. termékből eladott mennyiség a tárgyidőszakban
- $p_{0,j}$ ($p_{1,j}$): az j. termék egységára a bázis- (tárgy)időszakban
- $v_{0,j}$: a j. termék értékesítéséből származó árbevétel (tágabb értelemben *termelési érték*) a bázisidőszakban, számítása: $v_{0,j} = q_{0,j} \cdot p_{0,j}$
- $v_{1,j}$: a j. termék értékesítéséből származó árbevétel a tárgyidőszakban, számítása: $v_{1,j} = q_{1,j} \cdot p_{1,j}$

Érték-, ár- és volumenindexek II

Egyedi indexek – mostantól a j indexeket lehagyjuk

- Egyedi volumenindexek: $i_{q,j} = \frac{q_{1,j}}{q_{0,j}} \rightsquigarrow i_q = \frac{q_1}{q_0}$
- Egyedi árindexek: $i_{p,j} = \frac{p_{1,j}}{p_{0,j}} \rightsquigarrow i_p = \frac{p_1}{p_0}$
- Egyedi értékindexek: $i_{v,j} = \frac{v_{1,j}}{v_{0,j}} = \frac{q_{1,j} \cdot p_{1,j}}{q_{0,j} \cdot p_{0,j}} \rightsquigarrow i_v = \frac{v_1}{v_0} = \frac{q_1 p_1}{q_0 p_0} = i_p \cdot i_q$

Összetett indexek:

Index fajtája	Bázisidőszaki súlyozású vagy Laspeyres-féle	Tárgyidőszaki súlyozású vagy Paasche-féle	Fisher-féle
Árindexek	$I_p^0 = \frac{\sum q_0 p_1}{\sum q_0 p_0}$	$I_p^1 = \frac{\sum q_1 p_1}{\sum q_1 p_0}$	$I_p^F = \sqrt{I_p^0 \cdot I_p^1}$
Volumenindexek	$I_q^0 = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	$I_q^1 = \frac{\sum q_1 p_1}{\sum q_0 p_1}$	$I_q^F = \sqrt{I_q^0 \cdot I_q^1}$
Értékindex	$I_v = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum_{j=1}^{m} q_{1,j} p_{1,j}}{\sum_{j=1}^{m} q_{0,j} p_{0,j}}$		

Érték-, ár- és volumenindexek III

Állítás: az előző fóliákon bevezetett jelölésekkel igazak a következő összefüggések (indexek különböző átlagformulái):

- $I_v = I_q^0 \cdot I_p^1 = I_q^1 \cdot I_p^0 = \frac{\sum q_0 p_0 \cdot i_v}{\sum q_0 p_0} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_v}}$
- $I_p^0 = \frac{\sum q_0 p_0 \cdot i_p}{\sum q_0 p_0} = \frac{\sum q_0 p_1}{\sum \frac{q_0 p_1}{i_p}}$
- $I_q^1 = \frac{\sum q_0 p_1 \cdot i_q}{\sum q_0 p_1} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_q}}$

Megjegyzés: az egyes összetett indexek a megfelelő saját egyedi indexeik súlyozott átlagai

Az indexek képleteiben lévő osztások helyett különbségeket is lehet képezni, ekkor az I és i helyett K-t és k-t írunk. Például

$k_q = q_1 - q_0$

$K_p^0 = \sum q_0 p_1 - \sum q_0 p_0.$

Érték-, ár- és volumenindexek IV

Gyakorlati alkalmazásaik:

- Értékindexek:
  - vállalatok árbevételének, forgalmának alakulása
  - export és import értékének változása
  - energiafelhasználás értékének változása
- Árindexek:
  - fogyasztói árindex – az infláció mérőszáma, a lakosság által vásárolt fogyasztási cikkek és szolgáltatások árainak átlagos változását fejezi ki
  - cserearányindex: egy ország által exportált termékek árindexe osztva az általa importált termékek árindexével (itt tehát árindexekből számítunk további indexeket)

E**5.)** Ágnes asszony háromféle, saját termelésű termékeket árul a Lehel téri piac őstermelői részlegén: tojást, krumplit és hagymát. Az 2021-es és a 2020-as áprilisi forgalmát szeretné összevetni egymással. Eladásairól a következőket jegyezte fel:

Termékfajta	A forgalom értéke 2021-ben (e Ft)	Az árak	Az eladott mennyiség
		alakulása, 2021/2020 (%)	
Tojás	280	95	110
Krumpli	80	110	160
Hagyma	20	120	120

a.) Számítsunk érték-, ár- és volumenindexet a kofa forgalmára vonatkozóan! Értelmezzük szövegesen az egyes indexeket!

b.) A forgalom értékének növekedéséből hány forint volt az ár- és a volumenváltozás hatása?

Feladatok

Szöveges értelmezések I_v-re, I_p^1-re és I_q^0-ra:
Ágnes asszony 2021 áprilisában 16,1%-kal több bevételre tett szert, mint 2020 áprilisában. Ez két részből tevődött össze: egyrészt az eladott termékek árai – 2021-es eladott mennyiségekkel számolva – átlagosan 1,1%-kal csökkentek, másrészt az értékesítés volumene – bázisévi árakkal számolva – 17,4%-kal bővült.

Szöveges értelmezések K_p^1-re és K_q^0-ra:
Az árak átlagos csökkenése – 2021-es eladott mennyiségekkel számolva – a forgalom értékének csökkenéséhez 4.131 Ft-tal járult hozzá. Az eladott mennyiségek átlagos növekedése – bázisévi árakkal számolva – a forgalom értékének növekedéséhez 56.845 Ft-tal járult hozzá.

Mennyiségi sorok elemzése I

- ha a mennyiségi ismérv diszkrét és az ismérvváltozatok száma "kevés", akkor **gyakorisági sort** készítünk:

Ismérvértékek	Gyakoriságok
x_1	f_1
$\vdots$	$\vdots$
x_k	f_k
Összesen	n

  - n: minta mérete
  - k: különböző ismérvértékek száma
  - f_i: hányszor fordul elő az i-edik ismérvérték ($i = 1, \ldots, k$)

Mennyiségi sorok elemzése I/a

- ha a mennyiségi ismérv folytonos vagy "sok" ismérvváltozat van, akkor **osztályközös gyakorisági sort** készítünk:

Ismérvértékek	Gyakoriságok
$x_{1,a}$ – $x_{1,f}$	f_1
$\vdots$ $\vdots$	$\vdots$
$x_{k,a}$ – $x_{k,f}$	f_k
Összesen	n

  - $x_{i,a}$: az i-edik osztályköz alsó határa
  - $x_{i,f}$: az i-edik osztályköz felső határa
  - Minden megfigyelés pontosan egy osztályba kerüljön!

Mennyiségi sorok elemzése II

- Hány osztályköz legyen?
- Mik legyenek az osztályközök?

Hüvelykujjszabály:

- Osztályközök száma: $k = \lfloor \log_2 n \rfloor$
- Azonos hosszúságú osztályközök, hosszuk: $h = \frac{x_{max} - x_{min}}{k}$

Jelölések (osztályközös) gyakorisági soroknál:

- $x_i = \frac{x_{i,a} + x_{i,f}}{2} \rightsquigarrow$ az i. osztályközép/ismérvérték
- $f_i \rightsquigarrow$ gyakoriság
- $g_i = \frac{f_i}{\sum_i f_i} \rightsquigarrow$ relatív gyak.
- $s_i = x_i \cdot f_i \rightsquigarrow$ értékösszeg
- $f'_i = \sum_{k=1}^{i} f_k \rightsquigarrow$ kumulált gyak.
- $g'_i = \sum_{k=1}^{i} g_k \rightsquigarrow$ kum. rel. gyak.
- $s'_i = \sum_{k=1}^{i} s_k \rightsquigarrow$ kum. értékösszeg
- $z_i = \frac{s_i}{\sum_i s_i} \rightsquigarrow$ relatív értékösszeg
- $z'_i = \sum_{k=1}^{i} z_k \rightsquigarrow$ kumulált relatív értékösszeg

Példa - magyarországi jövedelmek 2015-ben

Decilis jövede-lemosztály	Bruttó éves átlag-jövedelem (e Ft)	Személyek száma (e fő)	g_i (%)	g'_i (%)	s_i (Mrd Ft)	z_i (%)	z'_i (%)
1	366	965	10	10	353	3	3
2	652	967	10	20	630	5	7
3	886	970	10	30	859	6	13
4	1025	964	10	40	989	7	20
5	1210	971	10	50	1175	8	29
6	1368	964	10	60	1318	9	38
7	1489	969	10	70	1443	10	49
8	1771	965	10	80	1710	12	61
9	2164	967	10	90	2092	15	76
10	3470	967	10	100	3355	24	100
Összesen	1440	9669	100	–	13924	100	–

Megjegyzések:

- A KSH (Központi Statisztikai Hivatal) háztartásokra összegezte a jövedelmeket, majd számolt átlagjövedelmet, így az átlagkeresetek gyerekekre is vonatkoznak, pedig ők nyilván nem dolgoznak.
- Ezek csak a legális jövedelmek, nincs bennük becslés az illegális jövedelmekre.
- A táblázatban lévő számok kerekített értékek.

Forrás: `https://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_zhc014a.html?down=1634`

Mennyiségi sorok elemzése III

Koncentráció: a sokasághoz tartozó teljes értékösszeg jelentős része a sokaság kevés egységére összpontosul.
Legyen a sokaság n elemű, a különböző ismérvértékek $x_1, \ldots, x_k$, ezek gyakoriságai $f_1, \ldots, f_k$.

Átlagos abszolút eltérés: $m = \frac{1}{n^2} \sum_{i=1}^{k} \sum_{j=1}^{k} f_i f_j |x_i - x_j|$.

A koncentráció mutatószámai:

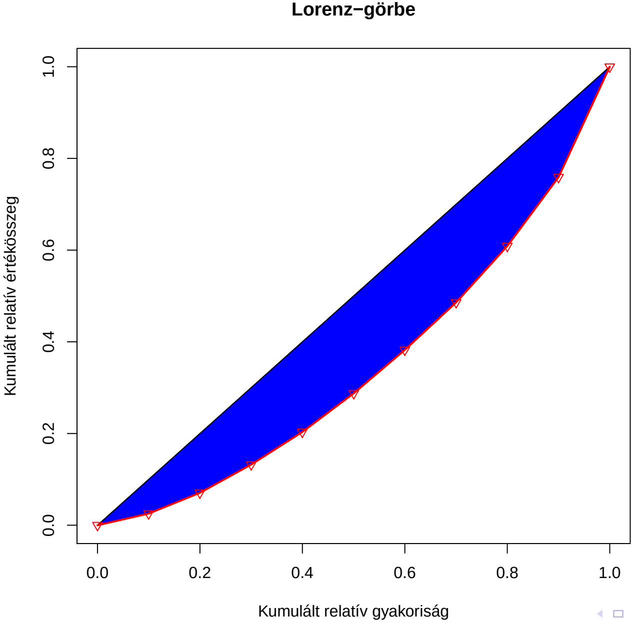
- *Gini-együttható (relatív abszolút eltérés fele)*: $G = \frac{m}{2\overline{x}}$
  Ez nem más, mint a koncentrációs terület (ld. következő fólia) 2-szerese.
  Értéke 0 és 1 között van; minél nagyobb, annál erősebb a koncentráció.
- *Herfindahl-index*: $HI = \sum_{i=1}^{k} z_i^2$, ahol z_i az i-dik csoport részaránya az összértékből ($\sum z_i = 1$)
  Értéke $\frac{1}{k}$ és 1 közötti; minél nagyobb, annál erősebb a koncentráció.

Mennyiségi sorok elemzése IV

Lorenz-görbe – a koncentráció mértékét szemléltető ábra

- Vízszintes tengely: g_i' kumulált relatív gyakoriságok
- Függőleges tengely: z_i' kumulált relatív értékösszegek
- A 45 fokos egyenes (átló) berajzolása
- *Koncentrációs görbe* berajzolása: $(0;0),(g_1';z_1'),(g_2';z_2'),\dots,(g_{k-1}';z_{k-1}'),(g_k';z_k')=(1;1)$ pontok összekötésével kapott töröttvonal
- *Koncentrációs terület*: a koncentrációs görbe és az átló által közbezárt terület
- Erős a koncentráció, ha a koncentrációs görbe közel van a négyzet oldalaihoz. Gyenge a koncentráció, ha a koncentrációs görbe közel van az átlóhoz.

Példa – magyarországi jövedelmek 2015-ben



Kék: koncentrációs terület

Piros: koncentrációs görbe (töröttvonal)

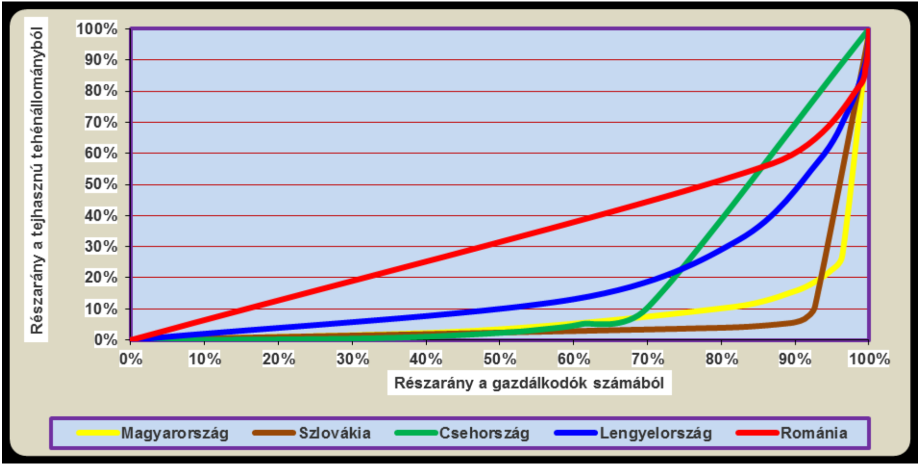
$G = 0,3089$

$HI = 0,134$
$0,1 \leq HI \leq 1$

Mihez viszonyítsuk a koncentráció mértékét?

Példa – Lorenz-görbe

A tejágazat koncentrációja különböző országokban



Forrás:
http://docplayer.hu/189090-A-magyar-tejagazat-helyzete-es-fejlodesenek-lehetseges-iranya.html

Feladatok

E**6.)** Egy piacon 4 azonos méretű vállalat működik (a piaci forgalomból azonos mértékben részesednek). Számszerűsítsük a Herfindahl-indexszel a piaci koncentráció változását, ha az egyik cég felvásárolja a másikat!

E**7.)** Számoljuk ki a csoport heti tanulási idejére a Lorenz görbét és a Gini index értékét!

E**8.)** Legyen az X valószínűségi változó

a.) eloszlása $P(X = 0) = P(X = 2) = P(X = 3) = \frac{1}{3}$;

b.) sűrűségfüggvénye $f(x) = (2x - 2)I(1 < x < 2)$.

Határozzuk meg X kvantilisfüggvényét!

E**9.)** Határozzuk meg a standard normális eloszlás móduszát, mediánját, ferdeségét és lapultságát!

E**10.)** Határozzuk meg a standard Cauchy-eloszlás (Cauchy(0;1)) és a Pareto-eloszlás várható értékét!

Eddig megismert eloszlások

Jelölése	Eloszlása		EX	D^2X
Ind(p)	$P(X=1)=p$	$P(X=0)=1-p$	p	$p(1-p)$
Hipgeo(N,M,n)	$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	$k=0,1,\ldots,\min(n,M)$	$n\frac{M}{N}$	$n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(1-\frac{n}{N}\right.$
Bin(n,p)	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	$k=0,1,\ldots,n$	np	$np(1-p)$
Geo(p)	$P(X=k)=p(1-p)^{k-1}$	$k=1,2,\ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
NegBin(n,p)	$P(X=k)=\binom{k-1}{n-1}p^n(1-p)^{k-n}$	$k=n,n+1,\ldots$	$\frac{n}{p}$	$\frac{n(1-p)}{p^2}$
Poi(λ)	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$	$k=0,1,\ldots$	λ	λ

Jelölése	Eloszlásfüggvény	Sűrűségfüggvény	EX	D^2X
$E(a,b)$	$\begin{cases} 0 & \text{ha } x\le a \\ \frac{x-a}{b-a} & \text{ha } a<x\le b \\ 1 & \text{ha } b<x \end{cases}$	$\begin{cases} \frac{1}{b-a} & \text{ha } a<x\le b \\ 0 & \text{különben} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$N(m,\sigma^2)$	$\ldots$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}\quad x\in\mathbb{R}$	m	σ^2
Exp(λ)	$\begin{cases} 1-e^{-\lambda x} & \text{ha } x\ge 0 \\ 0 & \text{különben} \end{cases}$	$\begin{cases} \lambda e^{-\lambda x} & \text{ha } x\ge 0 \\ 0 & \text{különben} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

További abszolút folytonos eloszlások

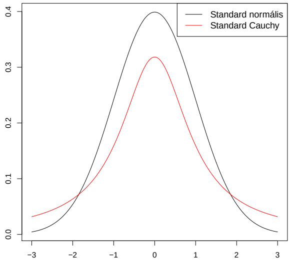
Eloszlás neve	Jelölése	Eloszlásfüggvény	Sűrűségfüggvény	EX	D^{2}X
Cauchy	$Cauchy(a, b)$ $a \in \mathbb{R}, b > 0$	$\frac{1}{\pi}\operatorname{arctg}\left(\frac{x-a}{b}\right) + \frac{1}{2}$	$\frac{1}{\pi b\left[1+\left(\frac{x-a}{b}\right)^2\right]}$ $x \in \mathbb{R}$	$\nexists$	$\nexists$
Pareto*	$Pareto(\alpha, \beta)$ $\alpha, \beta > 0$	$\begin{cases} 1-\left(\frac{\beta}{x}\right)^{\alpha} & \text{ha } x \geq \beta \\ 0 & \text{ha } x < \beta \end{cases}$	$\begin{cases} \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1} & \text{ha } x \geq \beta \\ 0 & \text{ha } x < \beta \end{cases}$	$\frac{\alpha\beta}{\alpha-1}$	$\frac{\beta^2\alpha}{(\alpha-1)^2(\alpha-2)}$

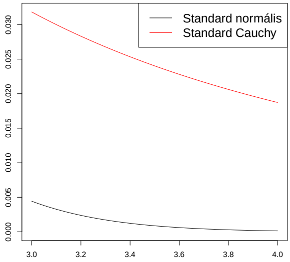
* A Pareto-eloszlásnak akkor van véges várható értéke a képletnek megfelelően, ha $\alpha > 1$, szórásnégyzete pedig akkor, ha $\alpha > 2$.

Eloszlás neve	Jelölése	Sűrűségfüggvény	EX	D^{2}X
Lognormális	$LN(m, \sigma^2)$ $m \in \mathbb{R}, \sigma > 0$	$\begin{cases} \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\log x - m)^2}{2\sigma^2}} & \text{ha } x \leq 0 \\ 0 & \text{ha} x < 0 \end{cases}$	$e^{m+\sigma^2/2}$	$(e^{\sigma^2}-1)e^{2m+\sigma^2}$
Gamma	$\Gamma(\alpha, \lambda)$ $\alpha, \lambda > 0$	$\begin{cases} \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}e^{-\lambda x}x^{\alpha-1} & \text{ha } x \geq 0 \\ 0 & \text{ha } x < 0 \end{cases}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Béta	$Beta(\alpha, \beta)$ $\alpha, \beta > 0$	$\begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} & x \in [0, 1] \\ 0 & \text{különben} \end{cases}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Khí-négyzet	χ^2_k $k \in \mathbb{N}$	$\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}$ $x \in \mathbb{R}$	k	$2k$
Student	t_ν $\nu > 0$	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	0 (ha $\nu > 1$)	$\frac{\nu}{\nu-2}$ (ha $\nu > 2$)
Fisher	F_{d_1,d_2} $d_1, d_2 > 0$	$\frac{\Gamma\left(\frac{d_1+d_2}{2}\right)}{\Gamma\left(\frac{d_1}{2}\right)\Gamma\left(\frac{d_2}{2}\right)}\left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}}x^{\frac{d_1}{2}-1}\left(1+\frac{d_1}{d_2}x\right)^{-\frac{d_1+d_2}{2}}$	$\frac{d_2}{d_2-2}$ (ha $d_2 > 2$)	$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ (ha $d_2 > 2$)

Vastag szélű (fat tailed) eloszlások

A sűrűségfüggvényük:





- végtelen vagy nem létezik a szórás (vagy a magasabb rendű momentumok)
- fontos vastag szélű eloszlások:
  - Cauchy-eloszlás
  - Pareto-eloszlás
  - Student-féle *t*-eloszlás alacsony szabadságfok esetén
- "extrém" események által okozott károk, például
  - nagy természeti katasztrófák,
  - atomerőmű-katasztrófák,
  - globális pénzügyi válságok,
  - az Internet összeomlása, stb.

  mértékének becslésére jóval alkalmasabbak a normális eloszlásnál

Standard normális versus vastag szélű eloszlások

Legyenek $X \sim$N(0;1), $Y \sim$Cauchy(0;1), $U \sim$Pareto$\left(2; \frac{1}{2}\right)$ és $V \sim t_2$

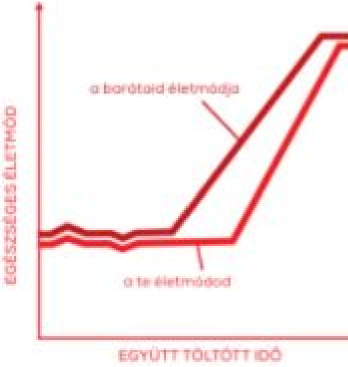
Ekkor a $h_i^Y = \frac{P(Y>i)}{P(X>i)}$, $h_i^U = \frac{P(U>i)}{P(X>i)}$ és $h_i^V = \frac{P(V>i)}{P(X>i)}$ hányadosok alakulása:

i	h_i^Y	h_i^U	h_i^V
2	$6,5$	$2,7$	$4,0$
3	$7,6 \cdot 10^1$	$2,1 \cdot 10^1$	$3,5 \cdot 10^1$
4	$2,5 \cdot 10^3$	$4,9 \cdot 10^2$	$9,0 \cdot 10^2$
5	$2,2 \cdot 10^5$	$3,4 \cdot 10^4$	$6,6 \cdot 10^4$
6	$5,3 \cdot 10^7$	$7,0 \cdot 10^6$	$1,4 \cdot 10^7$
7	$3,5 \cdot 10^{10}$	$4,0 \cdot 10^9$	$7,7 \cdot 10^9$
8	$5,9 \cdot 10^{13}$	$5,9 \cdot 10^{12}$	$1,1 \cdot 10^{13}$

Az eloszlások (paraméterfüggésük, kvantiliseik) itt is megnézhetőek:
http://www.randomservices.org/random/apps/SpecialCalculator.html

Rövid szünet

- Egy hír a bulvármédiából:
  - a balkezesek 7 évvel rövidebb ideig élnek
  - Ugyanis egy brit tudós megvizsgálta egy falu temetőjében található sírokat és mindenkinél utánajárt, hogy balkezes volt-e vagy sem
  - Mi lehet az igazság???
- Egy reklámkampány diagramjai



Középértékek számítása I

Adott az n elemű $\underline{x} = (x_1, x_2, \ldots, x_n)$ tapasztalati minta; osztályközös gyakorisági sor esetén k jelöli az osztályok számát, x_i az osztályközepeket, f_i pedig a gyakoriságokat.

Mintaátlag: az adatok átlagos értéke

- Számítása közvetlenül az adatokból: $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
- Számítása osztályközös gyakorisági sorból: $\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$

Módusz: a legtöbbször előforduló ismérvérték

- Számítása osztályközös gyakorisági sorból:

$$\text{Mo} = x_{mo,a} + \frac{d_a}{d_a + d_f} \cdot h_{mo}, \text{ ahol}$$

  - a móduszt tartalmazó osztályköz: amelyikben egységnyi osztályköz hosszra a legnagyobb gyakoriság jut ($\rightsquigarrow$ *korrigált gyakoriságok!*)
  - $x_{mo,a}$: a móduszt tartalmazó osztályköz alsó értéke
  - h_{mo}: a móduszt tartalmazó osztályköz hossza
  - d_a: a móduszt tartalmazó osztályköz korrigált gyakorisága mínusz a móduszt közvetlenül megelőző osztályköz korrigált gyakorisága
  - d_f: a móduszt tartalmazó osztályköz korrigált gyakorisága mínusz a móduszt közvetlenül követő osztályköz korrigált gyakorisága

Középértékek számítása II

Jelölje $x_1^* \leq x_2^* \leq \ldots \leq x_n^*$ a rendezett tapasztalati mintát.

Medián: azon ismérvérték, amelynél ugyanannyi kisebb vagy egyenlő, mint nagyobb vagy egyenlő ismérvérték fordul elő a mintában (a "középső" elem)

- Számítása közvetlenül az adatokból:

$$\mathrm{Me} = \begin{cases} x^*_{\frac{n+1}{2}}, & \text{ha } n \text{ páratlan} \\ \frac{x^*_{\frac{n}{2}} + x^*_{\frac{n}{2}+1}}{2}, & \text{ha } n \text{ páros} \end{cases}$$

- Számítása osztályközös gyakorisági sorból – két lépésben lineáris interpolációval:
  1. Melyik osztályközben van a medián: azon i, amire $f'_{i-1} \leq \frac{n}{2}$ és $f'_i \geq \frac{n}{2}$
  2. $\mathrm{Me} = x_{i,a} + \frac{\frac{n}{2} - f'_{i-1}}{f_i} \cdot h_i$, ahol
     - $x_{i,a}$: a mediánt tartalmazó osztályköz alsó értéke
     - h_i: a mediánt tartalmazó osztályköz hossza
     - f'_{i-1}: a mediánt közvetlenül megelőző osztályköz kumulált gyakorisága
     - f_i: a mediánt tartalmazó osztályköz gyakorisága

Tapasztalati kvantilisek számítása

Tapasztalati y-kvantilis: azon ismérvérték, amelynél a mintaelemek y-ad része kisebb vagy egyenlő, míg $(1-y)$-ad része nagyobb vagy egyenlő, $0 < y < 1$
Számítása nem egyértelmű, mi a szokásos interpolációs módszert alkalmazzuk két lépésben:

1. hányadik mintaelem a keresett kvantilis $\rightsquigarrow$ sorszám: $s := (n+1)y$
2. lineáris interpolációval a kvantilis kiszámítása

- Számítása közvetlenül az adatokból
  1. Sorszám: $s = e + t$ (e: egészrész, t: törtrész)
  2. $q_y = x_e^* + t(x_{e+1}^* - x_e^*)$
- Számítása osztályközös gyakorisági sorból – két lépésben lineáris interpolációval:
  1. Melyik osztályközben van az s-edik elem: jelölje ezt i, azaz $f'_{i-1} \leq s$ és $f'_i \geq s$
  2. $q_y = x_{i,a} + \frac{s - f'_{i-1}}{f_i} h_i$, ahol
     $x_{i,a}$, h_i, f'_{i-1} és f_i ugyanazokat jelöli, mint az előbb, csak az adott y-kvantilisre vonatkozóan

Nevezetes kvantilisek, további mérőszámok

A szakirodalomban a tapasztalati és az elméleti értékek között nem tesznek különbséget, mindegyiket nagy betűvel írják (ami néha meglehetősen zavaró...). Jelölje q_y a tapasztalati y-kvantilist.

- tercilisek: $T_1 = q_{1/3}$, $T_2 = q_{2/3}$
- **kvartlisek**:
  - $\mathbf{Q_1} = q_{1/4}$ (alsó kvartilis)
  - $Q_2 = \mathbf{Me} = q_{2/4}$ (középső kvartilis vagy medián)
  - $\mathbf{Q_3} = q_{3/4}$ (felső kvartilis)
- kvintilisek: $K_1 = q_{1/5}$, $K_2 = q_{2/5}$, $K_3 = q_{3/5}$, $K_4 = q_{4/5}$
- decilisek: $D_i = q_{i/10}$, $i = 1, 2, \ldots, 9$
- percentilisek: $P_i = q_{i/100}$, $i = 1, 2, \ldots, 99$

Ferdeség (skewness): $E(X - m)^3/\sigma^3$
Csúcsosság (kurtosis): $E(X - m)^4/\sigma^4 - 3$

Tapasztalati eloszlás

- **Tapasztalati eloszlás**: minden megfigyeléshez azonos, $\frac{1}{n}$ súlyt rendelünk $\quad \Rightarrow \quad$ ez egy diszkrét eloszlás
- A mintaátlag éppen ennek a várható értéke
- A tapasztalati eloszlás eloszlásfüggvényét hívjuk **tapasztalati eloszlásfüggvény**nek, ami egy tiszta ugrófüggvény, értéke minden mintaelem helyén $\frac{1}{n}$ nagyságot ugrik felfelé.
  A tapasztalati eloszlásfüggvény az x helyen:

$$\frac{I(x_1 < x) + I(x_2 < x) + \ldots + I(x_n < x)}{n} = \frac{\sum_{i=1}^{n} I(x_i < x)}{n}$$

  Azt mutatja meg, hogy a mintaelemek hányad része kisebb x-nél.