Valószínűségszámítás, statisztika és valóság

Néhány egyszerű példa

Kói Tamás
Budapesti Műszaki és Gazdaságtudományi Egyetem
koitomi@math.bme.hu

BME Nyílt Nap
2014. november 21.

Matematikai modell

- Matematikai modellel közelítünk egyes valós problémákat
- Sokszor egyszerűsítünk, próbálunk a lényegre koncentrálni
- Számolunk a matematikai modellben
- Akkor jó a modell, ha az eredmények hasznosíthatóak
- Születhetnek meglepő eredmények ...

Születésnap paradoxon

23 fős osztályban mi annak a valószínűsége, hogy van legalább két olyan diák, akiknek a születésnapja ugyanarra a napra esik?

Születésnap paradoxon

A matematikai modellben feltesszük, hogy:

- 365 napos az év
- Az emberek az év 365 napján egyforma eséllyel születnek

Klasszikus valószínűségi mezővel modellezhetjük a problémát, ahol egy elemi esemény egy 23 hosszú 1 és 365 közötti számokból álló sorozat:

$$\overbrace{220, 112, \ldots, 26, 26}^{23}$$

Így az összes elemi esemény száma: 365^{23}

Születésnap paradoxon

$$\mathbf{P}(\text{Van születésnap egyezés}) =$$
$$= 1 - \mathbf{P}(\text{A 23 diák az év 23 különböző napján született}) =$$
$$= 1 - \frac{365 \cdot 364 \cdot 363 \cdots 344 \cdot 343}{365^{23}}$$

Születésnap paradoxon

$$\mathbf{P}(\text{Van születésnap egyezés}) =$$
$$= 1 - \mathbf{P}(\text{A 23 diák az év 23 különböző napján született}) =$$
$$= 1 - \frac{365 \cdot 364 \cdot 363 \cdots 344 \cdot 343}{365^{23}}$$

- Ez körülbelül: 0.5073

Születésnap paradoxon

$$\mathbf{P}(\text{Van születésnap egyezés}) =$$
$$= 1 - \mathbf{P}(\text{A 23 diák az év 23 különböző napján született}) =$$
$$= 1 - \frac{365 \cdot 364 \cdot 363 \cdots 344 \cdot 343}{365^{23}}$$

- Ez körülbelül: 0.5073
- Elsőre meglepő: 23 diák van, míg 365 lehetséges születésnap

Születésnap paradoxon

$$\mathbf{P}(\text{Van születésnap egyezés}) =$$
$$= 1 - \mathbf{P}(\text{A 23 diák az év 23 különböző napján született}) =$$
$$= 1 - \frac{365 \cdot 364 \cdot 363 \cdots 344 \cdot 343}{365^{23}}$$

- Ez körülbelül: 0.5073
- Elsőre meglepő: 23 diák van, míg 365 lehetséges születésnap
- Intuitív magyarázat: a párok számítanak, amiből elég sok van

Születésnap paradoxon

$$\mathbf{P}(\text{Van születésnap egyezés}) =$$
$$= 1 - \mathbf{P}(\text{A 23 diák az év 23 különböző napján született}) =$$
$$= 1 - \frac{365 \cdot 364 \cdot 363 \cdots 344 \cdot 343}{365^{23}}$$

- Ez körülbelül: 0.5073
- Elsőre meglepő: 23 diák van, míg 365 lehetséges születésnap
- Intuitív magyarázat: a párok számítanak, amiből elég sok van
- Ha 50 fős az osztály, akkor ez a valószínűség: 0.9704

A valószínűség jelentése

- Egy kísérletben legyen az A esemény valószínűsége $\mathbf{P}(A) = p$
- Végezzünk el azonos körülmények között a kísérletet n-szer
- Jelüljük n_A-val azt a számot ahányszor az A esemény bekövetkezett
- Ekkor $\frac{n_A}{n} \approx p$

Feltételes valószínűség

- Legyenek A és B események egy kísérlethez kötődően
- Végezzünk el azonos körülmények között a kísérletet n-szer
- Jelöljük n_B-val azt a számot ahányszor a B esemény bekövetkezett
- Jelöljük $n_{A\cap B}$-val azt a számot ahány kísérletben az A és a B is bekövetkezett
- Vegyük észre, hogy:

$$\frac{n_{A\cap B}}{n_B} = \frac{\frac{n_{A\cap B}}{n}}{\frac{n_B}{n}} \approx \frac{\mathbf{P}(A\cap B)}{\mathbf{P}(B)}$$

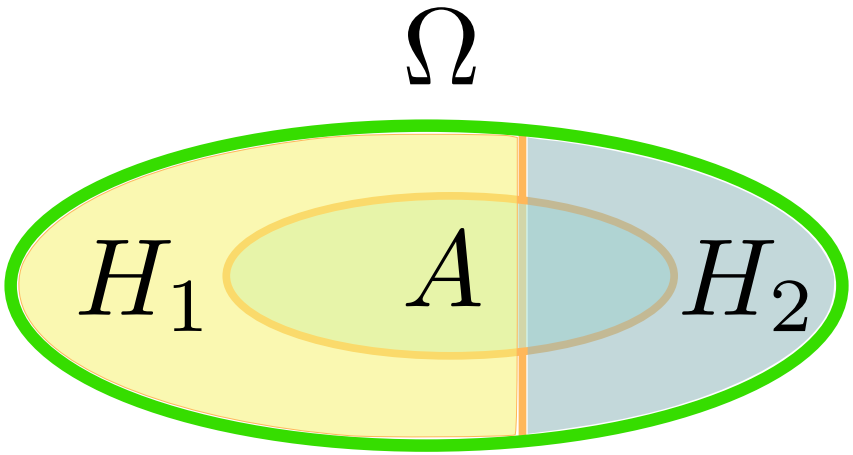
- A fenti miatt természetes a következő definíció:

$$\mathbf{P}(A|B) \triangleq \frac{\mathbf{P}(A\cap B)}{\mathbf{P}(B)}$$

- Átrendezett alak is fontos: $\mathbf{P}(A\cap B) = \mathbf{P}(B)\mathbf{P}(A|B)$

Bayes-tétel

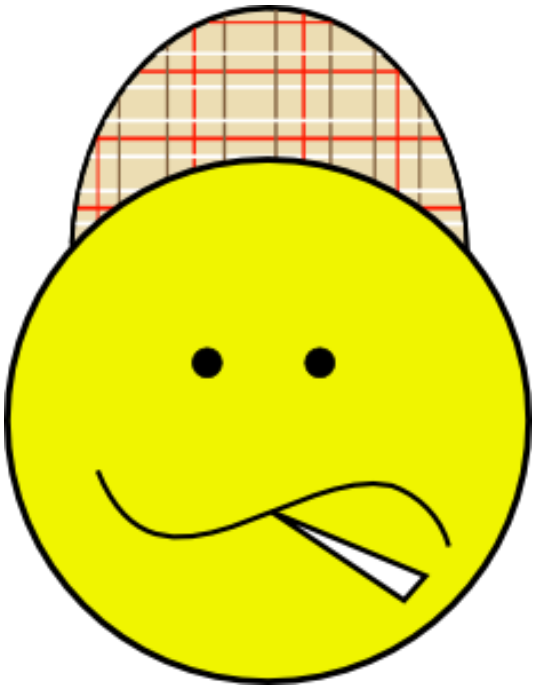
Legyen A egy tetszőleges esemény. Továbbá legyenek H_1 és H_2 olyan események, hogy bármilyen kimenetel esetén pontosan az egyikük következik be:



$$\mathbf{P}(H_1|A) = \frac{\mathbf{P}(H_1 \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(H_1 \cap A)}{\mathbf{P}(H_1 \cap A) + \mathbf{P}(H_2 \cap A)}$$

$$= \frac{\mathbf{P}(H_1)\mathbf{P}(A|H_1)}{\mathbf{P}(H_1)\mathbf{P}(A|H_1) + \mathbf{P}(H_2)\mathbf{P}(A|H_2)}$$

Beteg vagy nem beteg



Beteg vagy nem beteg

- A lakosságból véletlenül választunk egy egyént, akin elvégzünk egy bizonyos betegséghez kötődő újfajta szűrővizsgálatot
- H_1 és H_2 szerepében a "Beteg" illetve "Nem beteg" események vannak
- A szerepét a "Teszt pozitív" esemény tölti be
- Ismert a populációban a betegek aránya: $\mathbf{P}(\text{Beteg}) = 0.001$
- Ismert, hogy $\mathbf{P}(\text{Teszt pozitív}|\text{Beteg}) = 0.998$
- Ismert az is, hogy $\mathbf{P}(\text{Teszt pozitív}|\text{Nem beteg}) = 0.005$

$$\mathbf{P}(\text{Beteg}|\text{Teszt pozitív}) = \mathbf{P}(H_1|A)$$

$$= \frac{\mathbf{P}(H_1)\mathbf{P}(A|H_1)}{\mathbf{P}(H_1)\mathbf{P}(A|H_1) + \mathbf{P}(H_2)\mathbf{P}(A|H_2)} = \frac{0.001 \cdot 0.998}{0.001 \cdot 0.998 + 0.999 \cdot 0.005}$$

Beteg vagy nem beteg

- A lakosságból véletlenül választunk egy egyént, akin elvégzünk egy bizonyos betegséghez kötődő újfajta szűrővizsgálatot
- H_1 és H_2 szerepében a "Beteg" illetve "Nem beteg" események vannak
- A szerepét a "Teszt pozitív" esemény tölti be
- Ismert a populációban a betegek aránya: $\mathbf{P}(\text{Beteg}) = 0.001$
- Ismert, hogy $\mathbf{P}(\text{Teszt pozitív}|\text{Beteg}) = 0.998$
- Ismert az is, hogy $\mathbf{P}(\text{Teszt pozitív}|\text{Nem beteg}) = 0.005$

$$\mathbf{P}(\text{Beteg}|\text{Teszt pozitív}) = \mathbf{P}(H_1|A)$$

$$= \frac{\mathbf{P}(H_1)\mathbf{P}(A|H_1)}{\mathbf{P}(H_1)\mathbf{P}(A|H_1) + \mathbf{P}(H_2)\mathbf{P}(A|H_2)} = \frac{0.001 \cdot 0.998}{0.001 \cdot 0.998 + 0.999 \cdot 0.005}$$

- Utóbbi körülbelül 0.16

Simpson-paradoxon

Simpson-paradoxon

Kaliforniai Egyetem Posztgraduális felvételi adatai (1973)

	Felvételiző	Felvett
Férfi	8442	44%
Nő	4321	35%

Felmerül a nemi diszkrimináció vádja. Azonban:

Simpson-paradoxon

Kaliforniai Egyetem Posztgraduális felvételi adatai (1973)

	Felvételiző	Felvett
Férfi	8442	44%
Nő	4321	35%

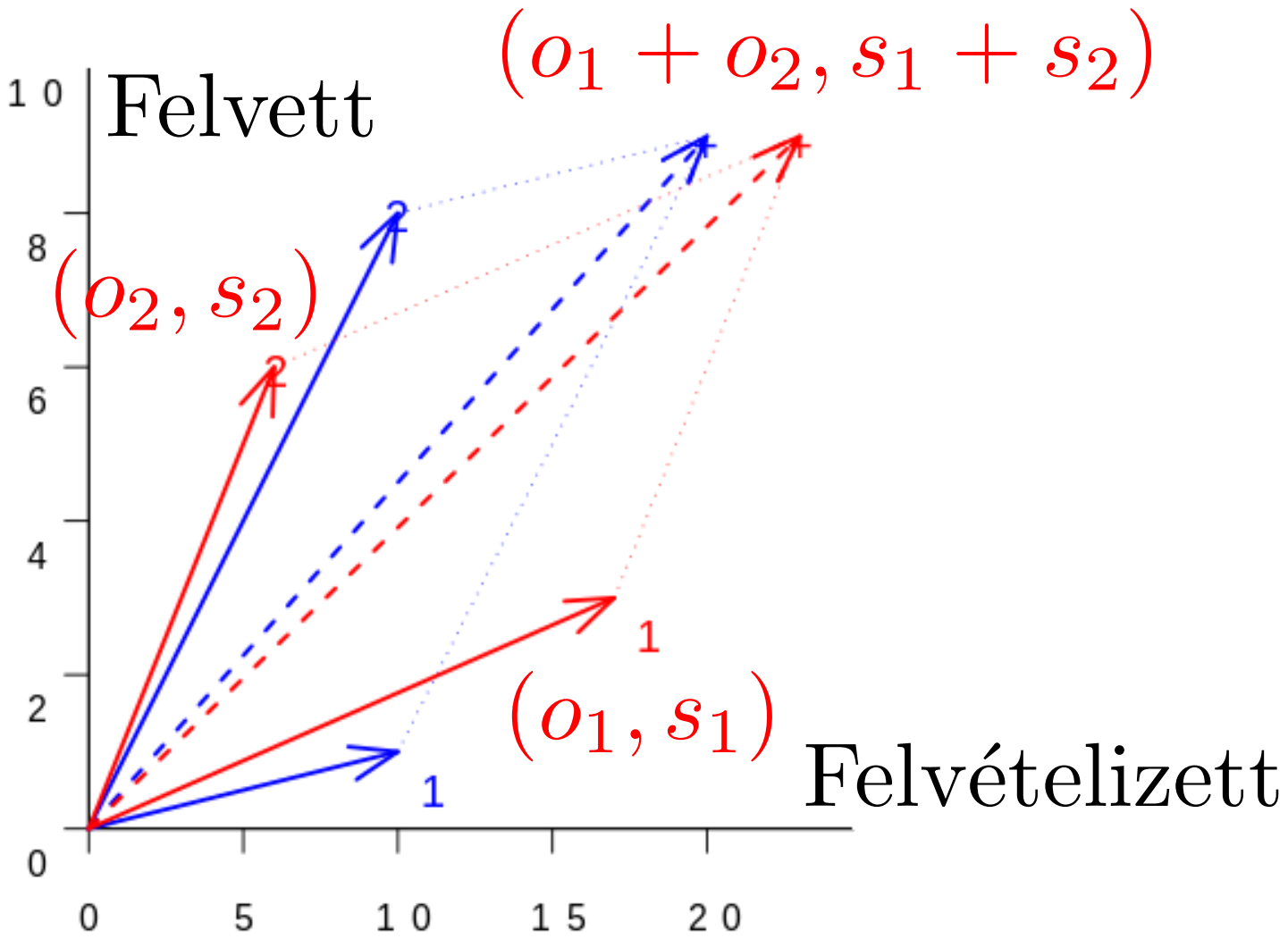
Felmerül a nemi diszkrimináció vádja. Azonban:

Kar	Férfi		Nő	
	Felvételiző	Felvett	Felvételiző	Felvett
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

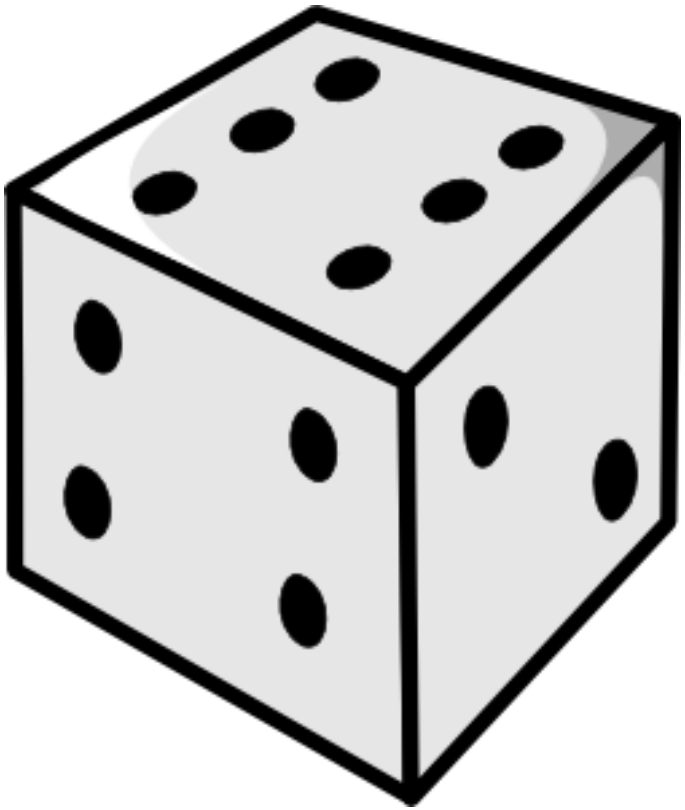
Simpson-paradoxon

- Matematikailag ez lehetséges
- Tényleg nagyobb arányban vették fel a férfiakat
- A probléma ott volt, amikor ok-okozati összefüggést feltételeztünk
- Nem a diszkriminálás miatt szerepeltek jobban a férfiak
- Hanem mert
  - A felvételi nehézsége karonként eltérő
  - A nők nagyobb arányban jelentkeztek a nehezebb szakokra
- Általánosabb nézőpontból összefüggést találtunk a "nem" és "sikeresség" változók között, ami a "kar" változó figyelembevételével eltűnt
- Konklúzió: legyünk körültekintőek a vádakat illetően ☺

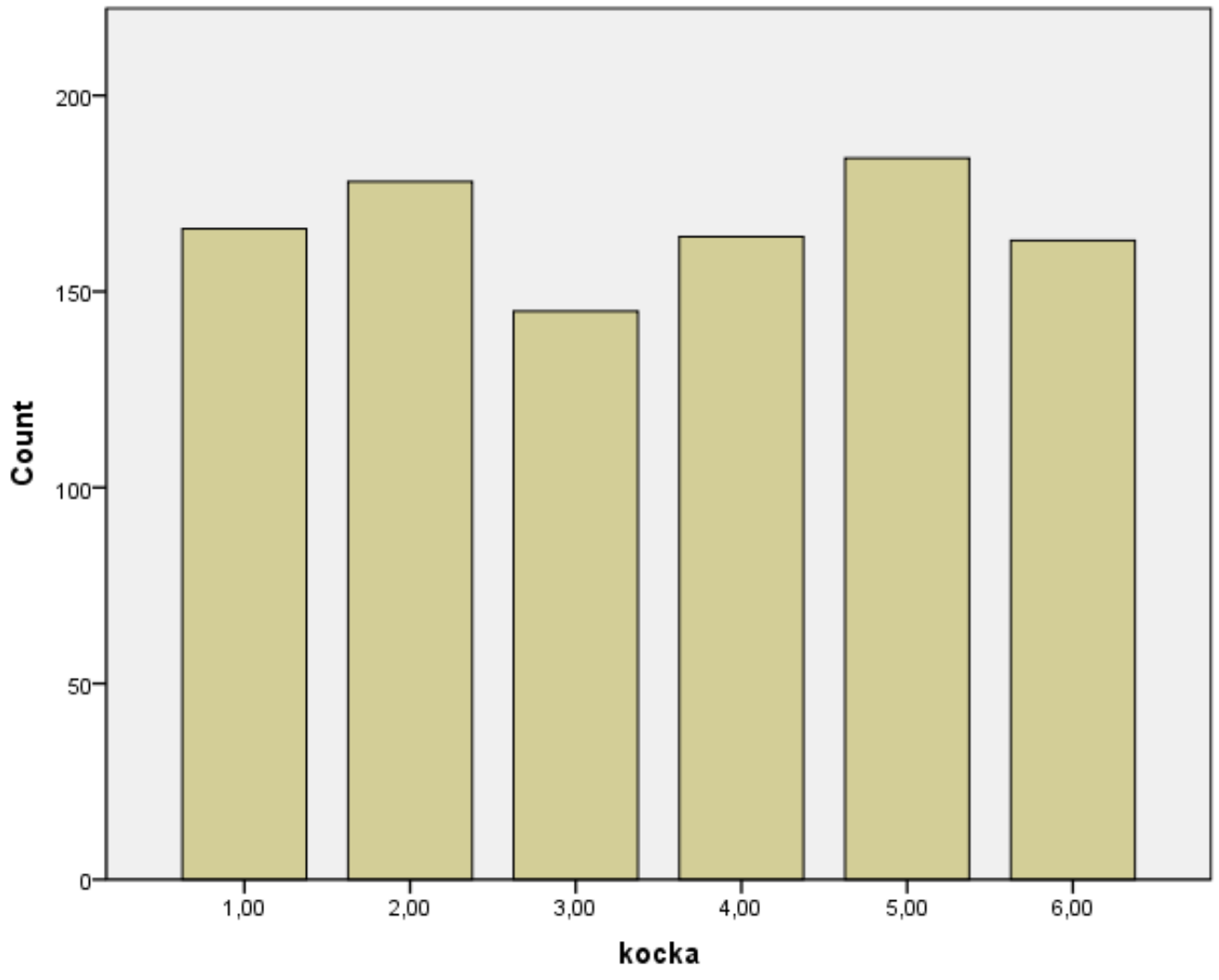
Simpson-paradoxon - matematikailag lehetséges



Cinkelt a kocka?



Cinkelt a kocka?



Cinkelt a kocka?

- A kérdés az, hogy amit látunk egyenletestől eltérést az indokolható-e véletlen fluktuációval, vagy arra kell gyanakodnunk, hogy cinkelt a kocka
- Alapelv az, hogy számolunk egy χ^2-el jelölt számot (statisztikát) a kapott adatokból
- χ^2 véletlen mennyiség: ha újra dobnánk sokat a kockával, akkor egy másik számot kapnánk
- Ha a kocka nem cinkelt, akkor ezt a véletlent tudjuk kezelni: (aszimptotikusan) ismerjük az eloszlását
- Tudjuk például hogy 0.95 a valószínűsége annak, hogy 11.1 alatt lesz ez a statisztika (elég sok dobás esetén)
- Azt is tudjuk, hogy ha a kocka cinkelt, akkor ez a statisztika a dobásszám növekedésével a végtelenhez tart

Cinkelt a kocka?

kocka

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	166	16,6	16,6	16,6
	2,00	178	17,8	17,8	34,4
	3,00	145	14,5	14,5	48,9
	4,00	164	16,4	16,4	65,3
	5,00	184	18,4	18,4	83,7
	6,00	163	16,3	16,3	100,0
	Total	1000	100,0	100,0	

$$\chi^2 = \frac{\left(166 - \frac{1000}{6}\right)^2}{\frac{1000}{6}} + \frac{\left(178 - \frac{1000}{6}\right)^2}{\frac{1000}{6}} + \cdots + \frac{\left(163 - \frac{1000}{6}\right)^2}{\frac{1000}{6}}$$

Ez körülbelül 5.51, ami kisebb, mint 11.1, így nincs matematikai okunk kételkedni abban, hogy a kocka nem cinkelt

Köszönöm a figyelmet!

- Vetier András angol nyelvű valószínűségszámítás elektronikus jegyzete: http://www.math.bme.hu/~vetier/df/index.html
- D. Freedman, R. Pisani, R. Purves, Statisztika, TYPOTEX, Budapest, 2005
- Wikipédia különböző fejezetei
- SPSS statisztika program
- Clip Art képek egyik forrása http://classroomclipart.com/
- Clip Art képek másik forrása http://www.clker.com/