

Amennyiben a statisztikai tábla készítésének célja a statisztikai sokaságot alkotó egyedek **osztályozása, csoportosítása**, akkor ezzel szemben három fontos követelményt kell támasztanunk.

1. Az osztályozás teljes legyen: a sokaság mindegyik egyedét be kell illesztenünk valamelyik osztályba.
2. Az osztályozás átfedés mentes legyen: a sokaság minden egyedét pontosan egy osztályba illeszthetjük be.
3. A csoportok homogének legyenek. Ez az elvárás a gyakorlatban sokszor nehezen teljesíthető.

Láthatjuk, hogy az osztályozásra, csoportosításra épülő sorok esetén a statisztikai sor első oszlopa tartalmazza az ismérvváltozatokat, míg a második oszlop valamilyen, a csoporthoz rendelt számértéket tartalmaz.

Osztály	Statisztikai jellemző
$C_1$	
$C_2$	
$C_k$	
<b>Összesen</b>	

**Fontos megjegyezni, hogy az összesen sor mindig olyan jellemzőt tartalmaz, ami az egész sokaságra vonatkozik, így az ide kerülő számérték – mint később látni fogjuk – nem mindig az oszlopban szereplő számok összegeként keletkezik.** Például, egy fiktív példát tekintve:

**Egy vállalat alkalmazottainak havi bruttó átlagkeresete**

Nem	Alkalmazottak száma, fő	Havi bruttó átlagkereset, ezer Ft
Férfi	30	270
Nő	20	250
<b>Összesen</b>	<b>50</b>	<b>262</b>

Az alkalmazottak száma oszlopban a számértékek összege megegyezik a vállalat alkalmazottainak számával, míg a havi bruttó átlagkereset esetében az összesen sorban a vállalat egészére vonatkozó átlagot, míg a felette lévő sorokban az egyes csoportokra vonatkozó adatot látjuk. Itt a 270 és a 250 összegének nincs tárgyi értelme.

A csoportokhoz rendelt számérték lehet az egyes osztályokba, csoportokba tartozó egyedek száma, melyet **gyakoriságnak** nevezünk. Az osztályok számát  $k$ , az  $i$ -edik osztály gyakoriságát  $f_i$  szimbólummal fogjuk jelölni, mely az angol frequency elnevezésből adódik. A gyakoriságok összege megadja a **sokaság elemszámát**, azaz a sokaság egyedeinek –  $N$  szimbólummal jelölt – számát:

$$\sum_{i=1}^k f_i = N.$$

**Egy lakótelepi panelház háztartásai taglétszám szerint**

Taglétszám (fő)	Háztartások száma (db)
1	5
2	10
3	7
4	25
5	2
6	1
<b>Összesen</b>	<b>50</b>

*Forrás: fiktív adatok*

A fenti példában  $f_1=5$ , azaz a panelházban 5 olyan háztartás van, melynek taglétszáma 1 fő.

Abban az esetben, ha az adott osztályozás esetében a sokaság megoszlására, azaz a százalékos összetételére vagyunk kíváncsiak, akkor kiszámíthatjuk az úgynevezett relatív gyakoriságokat (más néven **megoszlási viszonyszám**). Az  $i$ -edik osztály **relatív gyakorisága** megmutatja, hogy az adott osztályba a sokaság hány százaléka tartozik, melyet  $g_i$  szimbólummal fogunk jelölni. Ennek kiszámítása:

$$g_i = \frac{f_i}{\sum_{i=1}^k f_i} = \frac{f_i}{N}.$$

A relatív gyakoriság, mint hányados-képzésen alapuló összehasonlítás kifejezhető mind együtthatós, mind százalékos formában. Ekkor a relatív gyakoriságok összege 1, azaz 100% lesz.

Az alábbi táblázat a panelházi adattáblán szemlélteti a relatív gyakoriságok kiszámítását.

## Egy lakótelepi panelház háztartásainak taglétszám szerinti megoszlása

Taglétszám (fő)	Megoszlás (%)
1	$\frac{5}{50} = 0,1 \rightarrow 10$
2	$\frac{10}{50} = 0,2 \rightarrow 20$
3	$\frac{7}{50} = 0,14 \rightarrow 14$
4	$\frac{25}{50} = 0,5 \rightarrow 50$
5	$\frac{2}{50} = 0,04 \rightarrow 4$
6	$\frac{1}{50} = 0,02 \rightarrow 2$
<b>Összesen</b>	<b>100</b>

*Forrás: fiktív adatok*

Ekkor  $g_1=0,1$ ; azaz 10 százalék, azaz a panelházban háztartások 10 százalékának taglétszáma 1 fő.

A gyakoriságok és a relatív gyakoriságok az alábbi szerkezetben foglalhatóak táblázatba.

Osztály	Gyakoriság	Relatív gyakoriság
$C_1$	$f_1$	$g_1$
$C_2$	$f_2$	$g_2$
$C_k$	$f_k$	$g_k$
<b>Összesen</b>	<b>N</b>	<b>1 (vagy 100%)</b>

**Amikor csoportokhoz rendelünk számértékeket, akkor ezek ábrázolása történhet például oszlopdiagram segítségével. Gyakoriságokat, relatív gyakoriságokat emellett kördiagramon is megjeleníthetünk.**

Csoportosítás esetében a gyakoriságok, illetve a relatív gyakoriságok egymáshoz viszonyított arányát jellemezzük. Ezzel, a két csoport egymáshoz viszonyított arányát is meghatározhatjuk. Ezt néhány szakirodalom **koordinációs viszonyszám**nak is nevezi. Ennek meghatározása az alábbi képlettel történik.

$$\frac{f_i}{f_j} = \frac{g_i}{g_j}$$

Például az előző példán ez azt jelenti, hogy

$$\frac{25}{10} = \frac{50}{20} = \frac{0,5}{0,2} = 2,5$$

azaz, 2.5-szer annyi 4 fős háztartás van, mint 2 fős az adott panelban. Más megfogalmazásban egy kétfős háztartásra 2,5 négyfős háztartás jut. Mivel fél háztartás a gyakorlatban nem létezik, ezért ezt úgy szoktuk kifejezni, hogy 100 kétfős háztartásra 250 négyfős háztartás jut.

Egy-egy csoporthoz hozzárendelhetjük egy másik skálaváltozó ismérvváltozatainak összegét. A skálaváltozó ismérvváltozatait **ismérvértékeknek** is nevezzük és  $x_i$  szimbólummal jelöljük. Az  **$i$ -edik csoporthoz tartozó  $S_i$  értékösszeg** alatt az adott csoportba eső ismérvértékek összegét értjük, melyet az

$$S_i = \sum_{x_j \in C_i} x_j = f_i \cdot x_i$$

összefüggéssel határozhatunk meg. **Teljes értékösszeg** alatt – melyet  $S$  jelöl – az ismérvértékek összegét értjük, mely az osztályozások átfedés mentessége miatt kiszámítható a csoportokhoz tartozó értékösszegek összegeként is:

$$S = \sum_{j=1}^N x_j = \sum_{i=1}^k S_i = \sum_{i=1}^k f_i \cdot x_i = f_1 \cdot x_1 + f_2 \cdot x_2 + \dots + f_k \cdot x_k.$$

A táblázatok jelöléseiben ezentúl feltüntetésre kerülnek az adott mennyiségek statisztikai jelölései is. Ezek a táblák csak munkatáblaként használhatóak!

### Példák

**Egy vállalat alkalmazottainak havi bruttó keresetei**

Nem	Alkalmazottak száma, fő	Havi bruttó átlagkereset, ezer Ft	Havi bruttó keresetek összege, ezer Ft (Si)
Férfi	30	270	270*30=8100
Nő	20	250	20*250=5000
<b>Összesen</b>	<b>50</b>	<b>262</b>	<b>13100</b>

### Egy lakótelepi panelház lakói a háztartás taglétszáma szerint

Taglétszám, (fő) $x_i$	Háztartások száma (db) $f_i$	$S_i = f_i \cdot x_i$ (Lakók száma, fő)
1	5	5
2	10	20
3	7	21
4	25	100
5	2	10
6	1	6
<b>Összesen</b>	<b>50</b>	<b>162</b>

*Forrás: fiktív adatok*

Ekkor  $S_2=20$ , azaz 2 fős háztartásokban 20 fő él a vizsgált panelházban.  $S=162$ , mely megmutatja, hogy a vizsgált panelházban összesen 162 fő lakik. Az értékösszeg oszlop felirata úgy keletkezik, hogy végig gondoljuk, mit jelent a számítás, amit végrehajtunk.

Az  $i$ -edik csoporthoz tartozó  $Z_i$  relatív értékösszeg megmutatja, hogy az adott csoporthoz tartozó értékösszeg a teljes értékösszeg mekkora hányada:

$$Z_i = \frac{S_i}{S} = \frac{S_i}{\sum_{i=1}^k S_i} = \frac{f_i \cdot x_i}{\sum_{i=1}^k f_i \cdot x_i} = \frac{g_i \cdot x_i}{\sum_{i=1}^k g_i \cdot x_i}.$$

### Példa

#### Egy lakótelepi panelház lakóinak megoszlása a háztartás taglétszáma szerint

Taglétszám (fő)	$Z_i = \frac{S_i}{S}$ Lakók megoszlása (%)
1	$\frac{5}{162} = 0,0309 \rightarrow 3,09$
2	$\frac{20}{162} = 0,1235 \rightarrow 12,35$
3	$\frac{21}{162} = 0,1296 \rightarrow 12,96$
4	$\frac{100}{162} = 0,6173 \rightarrow 61,73$
5	$\frac{10}{162} = 0,0617 \rightarrow 6,17$
6	$\frac{6}{162} = 0,0370 \rightarrow 3,70$
<b>Összesen</b>	<b>100,00</b>

*Forrás: fiktív adatok*

Ekkor  $Z_2=0,1235$ , azaz 2 fős háztartásokban a vizsgált panelházban a lakók 12,35 százaléka él.

Az adatokat, elemi adatokat párhuzamosan **több szempont szerint** is megjeleníthetjük.

A statisztikai sokaságot alkotó egyedek **osztályozása, csoportosítása** párhuzamosan több szempont szerint is történhet, ilyenkor az úgynevezett **keresztáblához (kombinációs táblához)** jutunk. Például egy két szempontos keresztábla általános szerkezete az alábbi formában írható fel.

	<b>B1</b>	<b>B2</b>	...	<b>Bj</b>	...	<b>Bc</b>	<b>Összesen</b>
<b>A1</b>	$f_{11}$	$f_{12}$		$f_{1j}$		$f_{1c}$	$f_{1.}$
<b>A2</b>	$f_{21}$	$f_{22}$		$f_{2j}$		$f_{2c}$	$f_{2.}$
...							
<b>Ai</b>	$f_{i1}$	$f_{i2}$		$f_{ij}$		$f_{ic}$	$f_{i.}$
...							
<b>Ar</b>	$f_{r1}$	$f_{r2}$		$f_{rj}$		$f_{rc}$	$f_{r.}$
<b>Összesen</b>	$f_{.1}$	$f_{.2}$		$f_{.j}$		$f_{.c}$	$N$

Ekkor minden egyes kategóriát két index azonosít. Például, a megfigyelés során  $f_{21}$  darab olyan elemünk van, mely párhuzamosan az A szempont szerint a második, a B szempont szerint az első kategóriába tartozik. Ebben az esetben az  $f_{ij}$  értékeket gyakoriságoknak nevezzük. Minden sorban és minden oszlopban kiszámíthatóak az összesen értékek. Ezeket **peremgyakoriságoknak** nevezzük. Az i-edik sor összegét  $f_{i.}$ , míg a j-edik oszlop összegét  $f_{.j}$  szimbólummal jelöljük.

### Példa keresztáblára

#### A Statisztika I. kurzus hallgatói nem és szak szerint, 2017 (fő)

Szak	nem		<b>Összese</b> <b>n</b>
	férfi	nő	
Gazdálkodás és menedzsment	32	30	62
Kereskedelem és marketing	21	39	60
Pénzügy és számvitel	24	54	78
<b>Összesen</b>	<b>77</b>	<b>123</b>	<b>200</b>

*forrás: saját megfigyelés*

Ekkor az általános jelölések szemléltetésére néhány példa az alábbi:

- $f_{12}=30$ , azaz az  $N=200$  hallgató között 30 gazdálkodás és menedzsment szakos hölgy található.
- $f_{3.}=78$ , azaz a kurzusra 78 pénzügy és számvitel szakos hallgató járt
- $f_{.1}=77$ , azaz a kurzust 77 férfi vette fel.

Felmerül a kérdés, hogy kereszttáblák esetében milyen viszonyításban lehet relatív gyakoriságokat, megoszlásokat számolni. Három nevezetes viszonyítást említhetünk.

1. Kiszámíthatjuk az egyes csoportok arányát az összes megfigyelthez viszonyítva.

Szak	nem		Összesen
	férfi	nő	
Gazdálkodás és menedzsment	$\frac{32}{200} = 0,16 \rightarrow 16,0$	$\frac{30}{200} = 0,15 \rightarrow 15,0$	$\frac{62}{200} = 0,31 \rightarrow 31,0$
Kereskedelem és marketing	$\frac{21}{200} = 0,105 \rightarrow 10,5$	$\frac{39}{200} = 0,195 \rightarrow 19,5$	$\frac{60}{200} = 0,3 \rightarrow 30,0$
Pénzügy és számvitel	$\frac{24}{200} = 0,12 \rightarrow 12,0$	$\frac{54}{200} = 0,27 \rightarrow 27,0$	$\frac{78}{200} = 0,39 \rightarrow 39,0$
<b>Összesen</b>	$\frac{77}{200} = 0,385 \rightarrow 38,5$	$\frac{123}{200} = 0,615 \rightarrow 61,5$	$\frac{200}{200} = 1,0 \rightarrow 100,0$

Ekkor a táblázatból kiolvashatjuk többek között azt is, hogy a kurzus hallgatóinak 31 százaléka gazdálkodás és menedzsment szakos, 38,5 százaléka férfi és 16 százaléka gazdálkodás és menedzsment szakos férfi.

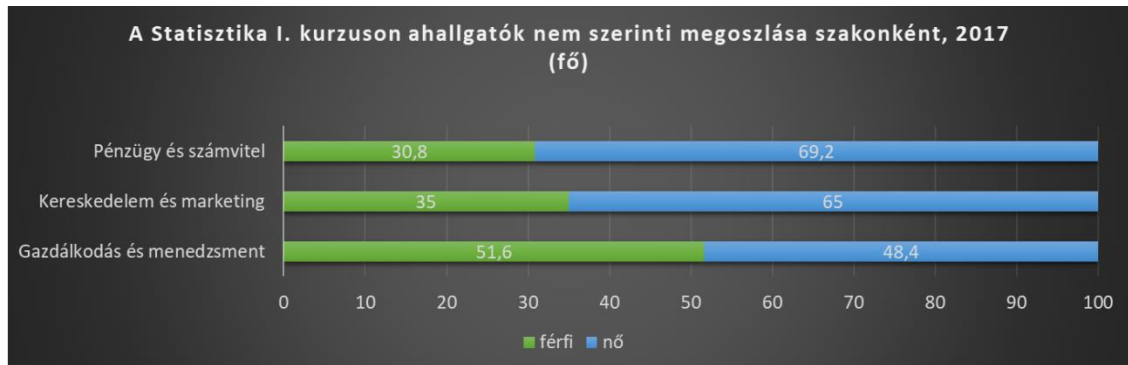
2. Kiszámíthatjuk az egyes csoportok arányát egy adott soron belül. Tehát mindenegyes gyakoriságot a saját sorának összegéhez viszonyítjuk.

Szak	nem		Összesen
	férfi	nő	
Gazdálkodás és menedzsment	$\frac{32}{62} = 0,516 \rightarrow 51,6$	$\frac{30}{62} = 0,484 \rightarrow 48,4$	$\frac{62}{62} = 1,0 \rightarrow 100,0$
Kereskedelem és marketing	$\frac{21}{60} = 0,35 \rightarrow 35,0$	$\frac{39}{60} = 0,65 \rightarrow 65,0$	$\frac{60}{60} = 1,0 \rightarrow 100,0$
Pénzügy és számvitel	$\frac{24}{78} = 0,308 \rightarrow 30,8$	$\frac{54}{78} = 0,692 \rightarrow 69,2$	$\frac{78}{78} = 1,0 \rightarrow 100,0$
<b>Összesen</b>	$\frac{77}{200} = 0,385 \rightarrow 38,5$	$\frac{123}{200} = 0,615 \rightarrow 61,5$	$\frac{200}{200} = 1,0 \rightarrow 100,0$

Ekkor gyakorlatilag minden szakra külön-külön meghatároztuk a nem szerinti megoszlást (összetételt). Tehát láthatjuk, hogy a gazdálkodás és menedzsment szakos hallgatók 51,6 százaléka férfi, 48,4 százaléka nő. Ez a számítás leegyszerűsítve a valószínűségszámításból tanult feltételes valószínűség empirikus megfelelője. (Milyen valószínűséggel férfi valaki, ha tudjuk, hogy az illető gazdálkodás és menedzsment szakos?)

A szöveges értelmezéseknél törekedjünk a pontos megfogalmazásokra. Például a kurzuson közel kétszer annyi férfi volt, mint ahány kereskedelem és marketing szakos hölgy, azonban a hallgatók között a férfiak aránya (38,5%) nem a kétszerese a nők

arányának a kereskedelem és marketing szakosokon belül (65%)! Sorszázalék esetében grafikus megjelenítésként halmozott sávdigramot is használhatunk.



3. Kiszámíthatjuk az egyes csoportok arányát egy adott oszlopon belül. Tehát mindenegyes gyakoriságot a saját oszlopának összegéhez viszonyítjuk.

Szak	nem		Összesen
	férfi	nő	
Gazdálkodás és menedzsment	$\frac{32}{77} = 0,416 \rightarrow 41,6$	$\frac{30}{123} = 0,244 \rightarrow 24,4$	$\frac{62}{200} = 0,31 \rightarrow 31,0$
Kereskedelem és marketing	$\frac{21}{77} = 0,273 \rightarrow 27,3$	$\frac{39}{123} = 0,317 \rightarrow 31,7$	$\frac{60}{200} = 0,3 \rightarrow 30,0$
Pénzügy és számvitel	$\frac{24}{77} = 0,311 \rightarrow 31,1$	$\frac{54}{123} = 0,439 \rightarrow 43,9$	$\frac{78}{200} = 0,39 \rightarrow 39,0$
<b>Összesen</b>	$\frac{77}{77} = 1,0 \rightarrow 100,0$	$\frac{123}{123} = 1,0 \rightarrow 100,0$	$\frac{200}{200} = 1,0 \rightarrow 100,0$

Ekkor gyakorlatilag mindkét nemre külön-külön meghatároztuk a szak szerinti megoszlást (összetételt). Tehát láthatjuk, hogy a férfiak 41,6 százaléka gazdálkodás és menedzsment, 27,3 százaléka kereskedelem és marketing, míg 31,1 százaléka pénzügy és számvitel szakos. Ez a számítás leegyszerűsítve a valószínűség számításból tanult feltételes valószínűség empirikus megfelelője. (Milyen valószínűséggel gazdálkodás és menedzsment szakos valaki, ha tudjuk, hogy az illető férfi?)

A szöveges értelmezéseknél törekedjünk a pontos megfogalmazásokra. Például a kurzuson a férfiak 41,6 százaléka gazdálkodás és menedzsment, a nők 43,9 százaléka pénzügy és számvitel szakos, azaz közel azonos, azonban az előbbi csoportba 32, míg az utóbbiba 54 fő tartozik!



SZEGEDI TUDOMÁNYEGYETEM  
GAZDASÁGTUDOMÁNYI KAR  
KÖZGAZDÁSZ KÉPZÉS  
TÁVOKTATÁSI TAGOZAT  
LECKESOROZAT  
COPYRIGHT © SZTE GTK 2017/2018

A LECKE TARTALMA, ILLETVE ALKOTÓ ELEMEI ELŐZETES,  
ÍRÁSBELI ENGEDÉLY MELLETT HASZNÁLHATÓK FEL.

JELEN TANANYAG  
A SZEGEDI TUDOMÁNYEGYETEMEN KÉSZÜLT  
AZ EURÓPAI UNIÓ TÁMOGATÁSÁVAL.  
PROJEKT AZONOSÍTÓ: EFOP-3.4.3-16-2016-00014

SZÉCHENYI 2020



MAGYARORSZÁG  
KORMÁNYA

Európai Unió  
Európai Szociális  
Alap



BEFEKTETÉS A JÖVŐBE