

**SÁNDORNÉ DR. KRISZT ÉVA – DR. CSESZNÁK ANITA
– ORSZÁG GÁBORNÉ**

Statisztika I.

Szerkesztette
SÁNDORNÉ DR. KRISZT ÉVA

Nemzedékek Tudása Tankönyvkiadó, Budapest

TARTALOMJEGYZÉK

ELŐSZÓ	9
1. A STATISZTIKA ALAPFOGALMAI	11
1.1. A statisztika tárgya és szerepe	11
1.2. A statisztikai egység, sokaság és ismerv	13
Mérési szintek	16
1.3. Statisztikai adat	17
1.4. Statisztikai csoportosítás és összehasonlítás	22
Csoportosítás	22
Összehasonlítás	24
1.5. Viszonyszámok	25
1.6. Átlagok	28
Számítási átlag	28
Harmonikus átlag	32
Mértani átlag	33
Négyzetes átlag	33
1.7. Az időbeli ismerv szerinti elemzés	35
1.7.1. Idősorok	35
1.7.2. Dinamikus viszonzyszámok	36
A bázis- és láncviszonyszámok közötti összefüggés	39
1.7.3. Az idősorok grafikus ábrázolása	40
1.7.4. Az idősorok elemzése átlagokkal	41
Az idősorok átlagos értékének meghatározása	41
Az idősor átlagos változásának vizsgálata	43
1.8. Gyakorlófeladatok	47
2. A MENNYISÉGI ISMÉRV SZERINTI ELEMZÉS	54
2.1. A mennyiségi ismerv	54
2.2. Gyakorisági sorok	56
Kumulált gyakorisági sorok	62
2.3. Értékösszegsor	64

2.4. A gyakorisági sorok grafikus ábrázolása	69
2.5. Helyzetmutatók	74
Módusz	74
Medián	78
Az átlag	80
Kvantilisek	84
2.6. Szóródási mutatók	88
A szóródás terjedelme	90
Átlagos eltérés	90
Szórás	92
Átlagos különbség	95
Relatív szórás	96
2.7. Az aszimmetria mérőszámai	97
2.8. A koncentráció elemzése	100
2.9. Gyakorlófeladatok	103

3. A SOKASÁGOK TÖBB ISMÉRV SZERINTI VIZSGÁLATA, A STATISZTIKAI TÁBLÁK ELEMZÉSE 107

3.1. A statisztikai táblákról általában	107
Az egyszerű táblák	108
A csoportosító táblák	109
A kombinációs táblák	109
3.2. Az egyszerű táblák elemzése	111
3.2.1. Intenzitási viszonyszámok és dinamikus viszonyszámok együttes alkalmazása	111
Intenzitási viszonyszámok dinamikus viszonyszámának számítása	116
3.2.2. A fejlődési tendenciák kimutatása, összehasonlítása	117
3.3. A csoportosító táblák elemzése	122
3.3.1. Rész- és összetett viszonyszámok	123
Az átlagos színvonal vizsgálata csoportosított sokaságok esetén	124
3.3.2. Szerkezet- és dinamikai változás vizsgálata	126
3.4. A kombinációs táblák (a sztochasztikus kapcsolatok) elemzése	132
3.4.1. Függvényszerű kapcsolat. Függetlenség	137
3.4.2. Az asszociáció szorosságának mérése	139

A Yule-féle asszociációs együttható	139
A Csuprov-féle és a Cramer-féle asszociációs együttható	143
3.4.3. A vegyes kapcsolat elemzése	147
Rész- és főátlagok	150
A rész- és főszakaságok szórása, szórásnégyzete	152
A vegyes kapcsolat szorosságának mérése	158
3.4.4. A korreláció elemzése csoportosított adatokból kiindulva	160
A korreláció szorosságának mérése	164
3.5. Gyakorlófeladatok	167

4. ÖSSZETETT INTENZITÁSI VISZONYSZÁMOK (FŐÁTLAGOK) ÖSSZEHASONLÍTÁSA 175

4.1. A standardizálás módszerének történeti áttekintése, a módszer lényege, jelölések	175
4.2. Az összetett intenzitási viszonyszámok (főátlagok) különbségének felbontása összetevőire	177
4.2.1. A részviszonyszámok (részátlagok) különbözőségének hatása . . .	180
4.2.2. Az összetétel különbözőségének hatása	182
4.3. Indexszámítás standardizálás alapján (hányadosfelbontás)	185
4.3.1. A részátlagindex	188
4.3.2. Az összetételhatás indexe	190
4.4. Alkalmazási területek	192
4.4.1. Az átlagbérek időbeli változásának vizsgálata	192
4.4.2. Az átlagárak időbeli változásának vizsgálata	197
4.5. Gyakorlófeladatok	201

5. ÉRTÉK-, ÁR- ÉS VOLUMENINDEXEK 207

5.1. Az indexszám fogalma	207
5.2. Érték-, ár- és volumenindex-számítás	208
5.2.1. Indexszám számítása aggregát formában	209
5.2.2. Az indexek átlagformái	212
Az értékindex átlagformái	212
Az árindex átlagformái	213
A volumenindex átlagformái	214
5.3. Az indexek súlyozása	217

5.4. Összefüggések az indexszámításban	220
5.4.1. Az indexszámok közötti összefüggések	220
5.4.2. Az aggregátumok közötti összefüggések	221
5.4.3. Csoportosított sokaságra számított indexek	223
5.5. Az indexszámok gyakorlati alkalmazása	224
5.6. Indexsorok	227
5.6.1. Az indexsorok közötti összefüggések	231
5.7. Területi indexek	233
5.8. Gyakorlófeladatok	236

ELŐSZÓ

Ez a tankönyv elsősorban a gazdaságtudományok képzési területéhez, ezen belül különösen az üzleti képzéshez tartozó felsőoktatási alapszakok első félévének leíró statisztikai módszertanát tartalmazza. Módszertani alaptankönyvként jól használható a társadalomtudományi és a gazdasági informatikus képzések esetében is.

A Budapesti Gazdasági Főiskolán korábban is használt módszertani alaptankönyv első kötetének továbbfejlesztett változata. Elsősorban azoknak ajánljuk, akik nem rendelkeznek előismeretekkel a statisztikai módszertanból. A számok világának megismerése, a statisztika nyelvének elsajátítása azonban elengedhetetlen az üzleti tanulmányok folytatásához, de megítélésünk szerint a mindennapokban való eligazodáshoz is.

A tankönyv a módszertani alapozást szolgálja. Az oktatásban erre épül a második féléves matematikai statisztikai tananyag, amely már feltételezi a valószínűség-számítási alapismereteket is. Az ehhez szükséges módszertani ismereteket a tankönyv második kötete tartalmazza.

A könyvben szereplő példák elsősorban a tananyag jobb megértését célozzák, de ugyanakkor törekedtünk a szemléletformálásra, a valóság, a társadalmi és gazdasági környezetünk megismerésének segítésére is. A feladatokat az elméleti előadásokhoz kapcsolódó gyakorlatokon kínáljuk a hallgatóknak. A megoldások során bemutatjuk a számítástechnikai alkalmazásokat, és hangsúlyt helyezünk az elemzésekre is.

A tankönyvet kiegészíti a példatár, amely számos megoldandó feladatot tartalmaz, segítve ezzel a begyakorlást és a tananyag készségszintű elsajátítását, hiszen a statisztikai módszertanban való kellő jártasság kialakítása a fő célunk. A példatár felépítése igazodik a tankönyvhöz, de több összefüggő feladatot, esettanulmányt is találnak abban a felhasználók, amelyek összefoglalják és bemutatják a tematikusan tárgyalt módszerek együttes alkalmazását.

Végezetül köszönetünket fejezzük ki mindazoknak, akik segítették munkánkat. Különösen lektorunknak, Prof. Dr. Besenyei Lajosnak, aki hasznos tanácsaival és jobbító észrevételeivel hozzájárult ahhoz, hogy ez a tankönyv elkészüljön. Köszönet illeti a Nemzedékek Tudása Tankönyvkiadó Zrt. munkatársait is, akik rugalmassága és ugyanakkor rendkívül precíz, igényes munkája nélkül ez a kiadvány nem jöhetett volna létre.

Kívánjuk, hogy a felhasználók sikeresen forgassák ezt a művet, és találják meg benne a statisztikai módszertan szépségét, érdekességét és hasznosságát, valamint leljék örömüket a statisztika segítségével szűkebb és tágabb környezetünk jobb megismerésében.

I.

A STATISZTIKA ALAPFOGALMAI

1.1. A statisztika tárgya és szerepe

Környezetünk megismeréséhez, a társadalmi és gazdasági folyamatok követéséhez, a makro- és mikrogazdasági jelenségek elemzéséhez és értékeléséhez egyre több információra van szükségünk. A folyamatok befolyásolásához, az ok-okozati összefüggések feltárásához elengedhetetlen a korszerű módszertani eszközök ismerete, de nélkülözhetetlenek a mindennapi dolgainkban való eligazodáshoz, valamint a megalapozott döntések meghozatalához is. Az információk között kitüntetett szerepe van a *számszerű információknak*, mert ezek a másféle információknál tömörebbek és egyértelműbbek. A számszerű információk gyűjtésében, feldolgozásában, elemzésében és publikálásában fontos szerepe van a statisztikának.

A statisztikáról kétféle megközelítésben is beszélhetünk. Statisztika alatt mindenekelőtt olyan számok vagy adatok összességét értjük, amelyek hűen leírják a környezetet. Az ilyen jellegű statisztika felhasználási célja szerint különbözőképpen épülhet fel. Jellemzően szabályos táblában helyezik el a statisztikai adatokat. A legtöbb ember a statisztikának ezt az értelmezését ismeri.

A másik jelentése szerint a statisztika a tömegjelenségek vizsgálatára szolgáló módszerek összességét öleli fel. A statisztikai módszerek alkalmazásának segítségével juthatunk el a vizsgált statisztikai sokaságnak, illetve azok egyedeinek tömör, számszerű jellemzéséhez.

A statisztika a valóság tömör, számszerű jellemzésére szolgáló tudományos módszertan, illetve gyakorlati tevékenység.

A statisztikai tevékenység az emberiség fejlődése folyamán jóval korábban kialakult, mint a statisztika tudománya. A statisztikai tevékenység úgyszólván egyidős az állammal. Kezdetben az állam fenntartásához szükséges információk (pl. a fegyverforgatásra alkalmas férfiak száma, a termények mennyisége) gyűjtése és közlése volt a feladata. Egy-egy népcsoportban, később nemzetben akkoriban elsősorban olyan alapvető kérdésekre keresték a választ, hogy például hányan vagyunk?, miből élünk?. A statisztika mint tudomány kialakulása azonban csupán a kapitalizmus kifejlődésével vette kezdetét. Ekkor a népesség és a termelés koncentrációja következtében a korábbi egyszerű nyilvántartási formák már nem voltak alkalmasak az egyre sokoldalúbb statisztikai jellegű állami és társadalmi igények kielégítésére. A statisztikatudomány fokozatosan fejlődve

– amihez nagy lendületet adott a valószínűség-számítás kialakulása és tételeinek elterjedése – önálló módszertudománnyá vált. Eredményeit a társadalomtudományok mellett széles körben alkalmazzák a természettudomány különböző területein is. A statisztikának egyre nagyobb szerepe van a gazdasági döntések előkészítése, az üzleti problémák elemzése mellett például az orvosi és biológiai kérdések megválaszolásában, a meteorológiai előrejelzések elkészítésében, a pszichológiában, de a jogi, bűnügyi jelenségek elemzésében is.

A statisztika mint gyakorlati tevékenység a tömegesen előforduló jelenségek egyedeire vonatkozó információk gyűjtése, feldolgozása, elemzése, ennek alapján a vizsgált jelenség egészének tömör, számszerű jellemzése.

A **statisztika** másrészt az információk összegyűjtésének, leírásának, elemzésének, értékelésének és közlésének **tudományos módszertana**.

A fenti megfogalmazásban igen fontos a **tömeges** jelző. A statisztika mindig tömegesen (nagy számban) előforduló jelenségek vizsgálatával foglalkozik. A statisztika nyelve ma már a mindennapi kommunikáció része. Aki nem érti ezt a nyelvet, nem tud eligazodni a XXI. századi világban. Gondoljuk meg, milyen kérdésekre kereshetik a környezetük iránt érdeklődők a választ. Például: mely gazdaságok növekednek a leggyorsabban?, melyek a legélhetőbb városok?, hol élnek a legtovább? és hol hálnak meg a legkorábban?, mely országok adósságterhe a legnagyobb?, melyek a leginnovatívabb államok?. Ezek a tömegjelenségek igen sokfélék lehetnek, de közös bennük, hogy mindegyikre számszerűsített információval adhatunk egzakt választ. Statisztikai módszerekkel vizsgálhatjuk például egy ország népességét, egy áruház forgalmát, egy ország gépkocsialományát, az energiatermelését, a lakosság fogyasztását is.

A statisztikai módszerek között vannak egészen egyszerű eljárások és vannak bonyolultabb, matematikai-statisztikai módszerek. A statisztikai módszertanon belül megkülönböztetünk *leíró statisztikát* és *statisztikai következtetést*.

A **leíró statisztika** az információk összegyűjtését, összegzését, tömör, számszerű jellemzését szolgáló módszereket foglalja magában. Idesorolhatjuk az adatgyűjtést, az adatok ábrázolását, csoportosítását, az adatokkal végzett egyszerűbb aritmetikai műveleteket, az eredmények áttekinthető formában való megjelenítését. Leíró statisztikai módszereket alkalmazunk például akkor, ha valamely település háztartásait (tömegjelenség) megfigyeljük taglétszámuk, jövedelmük, kiadásaik, fogyasztási szokásaik szerint. A begyűjtött információkat rögzítjük, majd csoportosítjuk a háztartásokat jövedelem, taglétszám stb. szerint, kiszámíthatjuk a háztartások átlagos jövedelmét, átlagos rezsiköltségét stb. A csoportosított adatokat, eredményeket szemléletes módon (ábrákkal, táblázatos formában) megjelenítjük, közzétesszük.

A **statisztikai következtetést** akkor alkalmazzuk, ha a tömegjelenségek egyedeinek teljes körű megfigyelésére nincs lehetőség, vagy a teljes körű megfigyelés túl költséges – így gazdaságtalan – és időigényes. Ilyen esetben az egye-

dek egy szűkebb csoportját figyeljük meg. A viszonylag kisszámú egyedre vonatkozó információk és az azokból számított eredmények alapján következtetünk a tömegjelenség egészére, jellemzőire, tulajdonságaira.

Következtetési statisztikai módszereket alkalmazunk például a közvélemény-kutatásoknál, a forgalomba kerülő termékek minőségének ellenőrzésekor, a lakosok életkörülményeinek vizsgálatánál. További alkalmazással találkozhatunk a különböző tényezők közötti összefüggések vizsgálatánál. A lakosság jövedelme (vagy annak változása) például miként befolyásolja a tartós fogyasztási cikkekre fordított kiadási összegeket (vagy azok változását), vagy a különböző ráfordítások hogyan befolyásolják a termelés eredményességét.

Könyvünk mind a leíró statisztikai, mind a következtetési statisztikai módszerek közül a komoly matematikai apparátust nem igénylő, leggyakrabban használt elemzési eszközöket tárgyalja, és kitér a gyakorlati alkalmazásokra is.

A statisztikai módszertant más szempont szerint is csoportosíthatjuk. Megkülönböztetünk általános statisztikát és szakstatisztikát.

Az **általános statisztika** a statisztika általános elméleti kérdéseivel, a statisztikai vizsgálatok során alkalmazásra kerülő módszerekkel általánosságban foglalkozik, a **szakstatisztika** a társadalmi-gazdasági élet egy-egy területének statisztikai módszerekkel való vizsgálatát tárgyalja. Ilyen például a népességstatisztika, az idegenforgalmi statisztika.

1.2. A statisztikai egység, sokaság és ismérv

A statisztika – mint már említettük – mindig tömegesen előforduló jelenségek és az azokat alkotó egységek vizsgálatával foglalkozik.

A **statisztikai egység** a statisztikai vizsgálat tárgyát képező egyed, a vizsgált információ hordozója.

Meghatározása nagy körültekintést igényel. A felsőoktatási intézmények számbavételénél például megfigyelési egységek az egyes intézmények. A statisztikai egységek azonban nem mindig határozhatóak meg ilyen egyértelműen. Ha egy meghatározott útszakaszon a forgalom nagyságát az időegység alatt áthaladó járművek számával kívánjuk jellemezni, a statisztikai egységek a forgalomszámláló pontok. Minden esetben lényeges az is, hogy mikor és hol történik a megfigyelés. Minden statisztikai egységnek tárgyi, területi és időbeli szempontból egyaránt egyértelműen körülhatárolhatónak kell lennie.

A **statisztikai sokaság** a megfigyelés tárgyát képező egyedek összessége, halmaza. A sokaságot alkotó egyedeket – a halmaz elemeit – a sokaság egységeinek nevezzük.

Statisztikai sokaságot alkothatnak *élőlények*, például a magyar felsőoktatás hallgatói, a Magyarországra érkező külföldi turisták, az ország loállománya;

tárgyak, például az ország lakásállománya, a kórházakban használt röntgenkészülékek; *szervezetek*, például a magyar főiskolák, ipari vállalkozások; *képzett egységek*, például bruttó hazai termék, gyümölcsfogyasztás.

Az élőlényekből, tárgyakból, szervezetekből álló sokaságok egyértelműen elkülönülő egységekből állnak. Az ilyen sokaságokat **diszkrét sokaságoknak** nevezzük. A képzett egységekből álló, ún. **folytonos sokaságoknál** az egységeket önkényesen határozhatjuk meg. Például a sokaság egy egysége: 1 Mrd Ft GDP, 1 kg gyümölcsfogyasztás.

A statisztikai sokaságok abból a szempontból is különböznek egymástól, hogy csak egy időpontra vonatkozóan vagy csak időtartamra vonatkoztatva értelmezhetők. Például egy megye népessége a természetes szaporodás (fogyás)¹ és a vándorlási különbözet² miatt állandóan változhat. Ezért e sokaság csak egy adott időpontban (pl. 2012. december 31. 0 óra, 00 perckor)³ értelmezhető, „ragadható meg”. Ugyanakkor a MOL Magyar Olaj- és Gázipari Nyrt. termelése – mivel a termelés egy folyamat – időpontban nem, csak egy időtartamban (a termelés egy napon, egy hónapban, egy évben stb.) értelmezhető. Az időpontra vonatkoztatva értelmezhető sokaságokat **álló sokaságoknak**, az időtartamra vonatkoztatva értelmezhetőket pedig **mozgó sokaságoknak** nevezzük.

A sokaság tartalmazhat *véges* (a gépkocsiállomány egy adott időpontban és területen) és *végtelen* (azonos körülmények között tetszőlegesen sokszor megismételhető kísérletek eredményeinek halmaza) elemszámú egyedet. A társadalmi-gazdasági vizsgálatokat általában véges számú egyed megfigyelésével végezzük.

A sokaságok fajtáit szemlélteti az *1.1. táblázat*:

1.1. TÁBLÁZAT
Példák a sokaságokra

A sokaság		
megnevezése	egysége	fajtája
Zala megye személygépkocsi-állománya 2012. december 31-én	egy személygépkocsi	diszkrét, álló, véges
Lakossági pénzügyi megtakarítások	egymilliárd (millió stb.) Ft betétállomány	folytonos, álló, véges
2012-ben Magyarországon született gyermekek	egy gyermek	diszkrét, mozgó, véges
Magyarország 2012. évi sörfogyasztása	egy liter (hektoliter, üveg stb.) sörfogyasztás	folytonos, mozgó, véges

¹ Az élveszületések és halálozások különbsége.

² A megyébe letelepülő és elköltöző népesség különbsége.

³ Ezt az időpontot a megfigyelés eszmei időpontjának szokás nevezni.

Amennyiben a statisztika a sokaságot meg kívánja ismerni, azt az egyedeken keresztül azok tulajdonságaival tudja megtenni, ugyanis bármely sokaság az egységei tulajdonságainak felsorolásával jellemezhető.

Statisztikai ismerv a statisztikai sokaság egyedeit jellemző tulajdonság. Az ismerv lehetséges kimenetelei az **ismervváltozatok**.

Ismerv például a gépkocsik típusa, fogyasztása, gyártási helye; ipari vállalkozásoknál a foglalkoztatottak száma, a bruttó kibocsátás, a területi elhelyezkedés, a vállalkozás profilja.

Ismervváltozatok például a gépkocsik típusánál az Audi, az Opel stb., az ipari vállalkozások területi elhelyezkedésénél Baranya megye, Békés megye stb., a gépkocsi fogyasztása esetén pedig számadatok.

Ha az ismerv csak két változattal rendelkezik, **alternatív ismervnek** nevezük. Ilyen például a nem (változatai: férfi, nő). A kettőnél több változattal rendelkező ismérvek is átalakíthatók alternatív ismervvé. Például az aktív keresők évi jövedelme kettőnél több változattal rendelkező ismerv (elvileg annyi változata lehetséges, ahány aktív kereső van), alternatív ismervvé alakítva: legfeljebb 500 000 Ft, illetve 500 000 Ft-nál nagyobb évi jövedelemmel rendelkezők.

Egy adott sokaságra vonatkozóan beszélhetünk **közös** és **megkülönböztető ismérvekről**. Azokat az ismérveket, amelyek szerint a sokaság egységei egyformák (pl. amelyek a sokaságot definiálják), *közös ismérveknek* nevezzük. Azokat az ismérveket, amelyek szerint az egyedek különböznek egymástól (ezek alapján a sokaság részsokaságokra bontható), *megkülönböztető ismérveknek* nevezük.

Ha a megfigyelt sokaságot a Budapesti Gazdasági Főiskola Pénzügyi és Számviteli Kara nappali tagozatára 2012. szeptember 11-én beiratkozott I. évfolyamos hallgatók képezik, akkor a definiáló közös ismérvek: a beiratkozás helye (BGF PSZK), az évfolyam (I.), a beiratkozás időpontja (2012. szeptember 11.), megkülönböztető ismérvek például a hallgatók neme, iskolai végzettsége, lakcíme, életkora, a felvételi vizsgán elért pontszáma.

Attól függően, hogy az ismervváltozatok milyen jellegű információt adnak a sokaság egyedeiről, különböző fajta ismérveket különböztetünk meg. Az ismérvek fajtái:

- **időbeli ismérvek:** az egységek időbeli elhelyezésére szolgáló rendezőelvek. Ismervváltozatai: időpontok, időszakok.
(Pl. a főiskolai hallgatók születési ideje, a gépkocsik gyártási ideje.)
- **területi ismérvek:** az egységek térbeli elhelyezésére szolgáló rendezőelvek. Ismervváltozatai: földrajzi egységek.
(Pl. a főiskolai hallgatók lakhelye, a gépkocsik gyártási helye.)
- **minőségi ismérvek:** az egyedek számszerűen nem mérhető tulajdonságai.
(Pl. a főiskolai hallgatók neme, a gépkocsik típusa.)

- **mennyiségi ismérvek:** az egyedek számszerűen mérhető (megszámlálható) tulajdonságai. Ismérvváltozatait **ismérvértékeknek** nevezzük. (Pl. a főiskolai hallgatók felvételi pontszáma, a gépkocsik fogyasztása.)

Mérési szintek

A felsorolt ismérvek közül csak a mennyiségi ismérv változatai számadatok. Bizonyos szabályok betartása mellett azonban minden ismérv lehetséges változatai számértékekké alakíthatók. Például a nem ismérvének két változata van: férfi és nő. E változatokhoz számértékeket rendelhetünk: férfi: 1; nő: 2. Ilyen alapon a sokasági egységek bármilyen tulajdonságának megfigyelése és rögzítése az egységek számokkal való jellemzésének, azaz mérésének tekinthető.

A mérés számok meghatározott szabályok szerinti hozzárendelése jelenségekhez (dolgokhoz, tárgyakhoz, eseményekhez), illetve ezek bizonyos tulajdonságaihoz.

A hozzárendelési szabályok alapján négy mérési skálát (szintet) különböztetünk meg:

- névleges skála,
- sorrendi skála,
- intervallumskála,
- arányskála.

A **névleges (nominális) mérési skála** (mérési szint) a számok kötetlen hozzárendelését jelenti. Nominális skálát alkalmazunk a területi és minőségi ismérvek szerinti megfigyeléseknél. E skálán való méréskor a számok (kódszámok) csak a sokaság egyedeinek azonosítására szolgálnak. Ilyen nominális skála például a rendszám, irányítószám, biztosítási szám. Mivel a számok csak az egyedek azonosítását (megkülönböztetését) szolgálják, közöttük az egyéb relációk (pl. nagyobb, kisebb) nem értelmezhetők, ezért e számokkal végzett különböző számtani műveleteknek semmi értelme nincs.

A **sorrendi (ordinális) mérési skála** a sokaság egyedeinek egy közös tulajdonság alapján való sorba rendezése. Ilyen sorrendi skála például a hallgatók osztályzata, a tornászok helyezési sorrendje, az országok hitelképességi sorrendje, a termékek minőségi osztályai. A skálán – bár a sorszámok közötti különbség azonos (egy-egy) – az egyes egyedek nem feltétlenül egyenlő távolságra helyezkednek el egymástól. Az első és a második helyre sorolt tornász teljesítménye között például nem biztos, hogy ugyanakkora a különbség, mint a negyedik és ötödik helyezetteké között. A mérésből származó adatokkal (sorszámokkal) ezért csak azok a műveletek végezhetők el, amelyek során kizárólag a skálát képező számértékek sorrendisége kerül kihasználásra.

Az **intervallumskála (különbségi skála)** már a szó hagyományos értelmében is mérést jelent, ugyanis a skálaértékek különbségei is valós információt

adnak a sokaság egységeiről. Az intervallumskálának egy jellegzetes tulajdonsága, hogy a mértékegység és a nullapont meghatározása önkényes, és e nulla érték nem tükrözi a tulajdonság hiányát. Ilyen skálán mérik például a hőmérsékletet. Ha a skála mértékegysége a Celsius-fok (pl. a Fahrenheit-fok, illetve a Kelvin-fok is használatos mértékegység), a skála nullapontja a víz fagyáspontja, és ez nyilvánvalóan nem tekinthető abszolút nullapontnak. A skálán két érték összege vagy aránya nem értelmezhető. Például nem mondhatjuk, hogy a $+20\text{ °C}$ -os és a $+5\text{ °C}$ -os hőmérséklet összege $+25\text{ °C}$, vagy hogy a 20 °C -os hőmérséklet kétszerese a 10 °C -osnak. Ugyanakkor két-két adat különbsége, a két különbség összege és aránya már értelmezhető. Például az 5 °C és a 10 °C közötti különbség azonos a 15 °C és a 20 °C közötti különbséggel. A 20 °C és a 30 °C közötti különbség kétszerese az előbbi bármelyik különbségnek.

Az **arányskálán** történő mérés – a legmagasabb mérési szint – nyújtja a legtöbb információt. A skálának valódi nullapontja van, mely a tulajdonság hiányát jelzi. A skála bármely két értékének aránya független a mértékegységtől. E skálán nyert számokkal a statisztikai elemzésekhez szükséges összes művelet elvégezhető. Arányskálán mért értékek például a hosszúság, a jövedelem, a költség, a termelés mennyisége.

1.3. Statisztikai adat

A statisztikatudomány fontos alapfogalmai közé tartozik a statisztikai adat fogalma.

A statisztikai adat valamely statisztikai sokaság elemeinek száma vagy a sokaság valamilyen másféle számszerű jellemzője, mérési eredmény.

A statisztikai adat mindig tartalmaz fogalmi jegyeket, időbeli, térbeli vagy másféle azonosítókat (mértékegységet) és ezek mellett egy számértéket. A statisztikai adat tehát nem pusztán a számérték maga.

Statisztikai adat például: 2012-ben a magyar felsőoktatás költségvetési támogatása 158 Mrd Ft volt; Magyarország népessége 2011. október 1-jén a népszámlálás előzetes adatai szerint 9 millió 982 ezer fő volt; 2010-ről 2011-re a fogyasztói árindex 3,9%-kal növekedett.

Azokat a statisztikai adatokat, amelyekhez mérés vagy számlálás útján jutunk, **alapadatoknak** nevezzük (almatermés, népesség száma).

Két vagy több alapadattal végzett műveletek eredményeként **leszármaztatott adatokhoz** jutunk. Például Magyarországon 2010. december 31-én az összes foglalkoztatott száma 3781,2 ezer fő volt, 2011. december 31-én pedig 3811,9 ezer fő. E két alapadatból osztással képzett leszámaztatott adat:

$$\frac{3811,9}{3781,2} = 1,0081.$$

A foglalkoztatottak száma egy év alatt 1,0081-szeresére, vagyis 100,8%-ra, azaz 0,8%-kal nőtt.

Statisztikai mutatószámok: azok a statisztikai adatok (általában leszármaztatott adatok), amelyekkel valamilyen rendszeresen megisméltendő (pl. társadalmi, gazdasági) jelenséget statisztikailag jellemezhetünk.

Az életszínvonal egyik mutatószáma például az 1 főre jutó reáljövedelem, a gazdasági fejlettség az 1 főre jutó GDP, a termelékenység az 1 órára jutó termelés.

A statisztikai vizsgálatok kiindulópontját az alapadatok képezik. Az alapadatokkal szemben többféle követelményt támasztunk:

- Az adatok legyenek a felhasználás szempontjából elfogadható pontosságúak. Minél pontosabbak az adatok, annál megalapozottabb döntéseket hozhatunk.
- Az adatok kellő időben álljanak rendelkezésre. Az adatszolgáltatás gyorsasága ugyanis fontos szerepet játszik a társadalmi-gazdasági folyamatok alakításában.
- A szükséges adatokhoz a lehető legkisebb ráfordítással (költséggel) jussunk hozzá.

E követelményeknek – az **elfogadható pontosság**, a **gyorsaság** és a **gazdaságosság** – egy időben általában tökéletesen megfelelni nem lehet (pl. a gyorsaság a pontosság ellen hat).

Az alapadatokhoz többféle módon juthatunk. A statisztikai elemzések forrását képezhetik az eredetileg nem statisztikai célra készült kimutatások, nyilvántartások. (Pl. az önkormányzatok lakónyilvántartása, gépkocsik nyilvántartásai, a gazdasági szervezetek különféle számviteli nyilvántartásai.)

A statisztikai adatok másik forrását az e célra szervezett **adatgyűjtések (adatfelvételek)** képezik. Az adatgyűjtést (adatfelvételt) minden esetben megelőzi egy olyan adatfelvételi program kidolgozása, amelyben a statisztikai tevékenység egészét megtervezzük. Az adatfelvétel végrehajtása előtt a vizsgálat eredményessége szempontjából tisztázni kell a felvétel célját, az adatfeldolgozás, az elemzés és a közlés menetét. Ennek elmaradása használhatatlan alapadatokhoz, téves információhoz vezethet. Az ilyen adatok, információk pedig megalapozatlan, hibás döntéseket eredményezhetnek. Ez csak úgy kerülhető el, hogy a vizsgálat céljának alárendelten tervezzük meg a statisztikai tevékenység teljes folyamatát az adatgyűjtéstől kezdve az adatközlésig.

Az adatgyűjtés megtervezésénél dönteni kell arról is, hogy az adatfelvétel a vizsgált sokaság minden egységére kiterjedjen-e, vagy csak a sokaság megfelelő módon kiválasztott részére. Az adatfelvétel, attól függően, hogy a sokaság mekkora részére terjed ki, lehet *teljes körű* és *részleges*.

A **teljes körű felvétel** a vizsgált sokaság valamennyi egyedére kiterjed. Ilyen

felvételt csak véges elemszámú sokaság esetén lehet megvalósítani. A teljes körű megfigyelések jellegzetes példái a népszámlálások.

A **részleges felvétel** a sokaságnak csak egy kiválasztott részére terjed ki. Végtelen elemszámú sokaság megfigyelése csak részleges adatfelvétellel lehetséges. Véges és nagyszámú sokaság esetén is gyakran kerül sor azonban ilyen felvételekre. Ennek elsősorban az a magyarázata, hogy a sokaság teljes körű megfigyelése jelentős költséggel jár és időigényes. Egy szakszerűen végzett részleges megfigyelés, amellyel együtt olcsóbb és gyorsabb a teljes körű felvételnél, alkalmas a teljes sokaságra vonatkozó következtetések levonására is.

A részleges adatfelvétel jellegzetes típusai: a reprezentatív adatfelvétel, a monográfia és egyéb részleges (nem reprezentatív) adatfelvételek.

Reprezentatív (mintavételes) adatfelvételnek nevezzük a részleges felvételnek azt a fajtáját, amelynél a megfigyelésbe vont részsokaság kiválasztása meghatározott elvek, módszerek alapján történik, és a kiválasztott részsokaság hűen tükrözi (reprezentálja) az egész sokaságot. A megfigyelt sokaság egészét **alapsokaságnak**, a kiválasztott részsokaságot **mintasokaságnak** vagy röviden **mintának** nevezzük. A mintából származó minden eredményt a sokaság egészének jellemzésére használunk fel, a felvétel részlegessége ellenére a sokaság egészére általánosítjuk. E mintából való következtetés – éppen a felvétel részlegessége miatt – csak bizonyos hibával valósítható meg, amit **mintavételi hibának** nevezünk.

Ilyen mintavételes adatfelvételt alkalmazunk például a lakosság jövedelmének, fogyasztási szokásainak vizsgálatánál; a mezőgazdaságban a várható termésmennyiség becsléséhez; a közvélemény-kutatásoknál.

A **monográfia** a sokaság egy vagy néhány kiemelt egyedének részletes statisztikai vizsgálatát jelenti. Ilyen például egy nagyon jó és egy nagyon rossz eredményt elérő bank tevékenységének, gazdálkodásának sokoldalú elemzése.

Egyéb, részleges adatgyűjtéssel is találkozhatunk a gyakorlatban. Például ha egy adott termék (pl. egy mosópor) vásárlói kérdőívet kapnak, és az önként kitöltött és beküldött kérdőíveket feldolgozzák. Az ilyen adatgyűjtések, bár hasznos információkat szolgáltatnak, nem általánosíthatók az alapsokaságra.

A részleges felvétel megismert típusai közül a társadalmi-gazdasági statisztika legfontosabb és leggyakrabban használt módszere a reprezentatív megfigyelés.

Az adatgyűjtések során általában **kérdőíveket** használunk, melyek a kérdések mellett a válaszok rögzítésére szolgáló üres rovatokat is tartalmaznak. A kérdőív lehet **egyéni kérdőív** és **lajstrom**. Az egyéni kérdőívre egy, a lajstromra több **megfigyelési egység** adatai kerülnek. A felvétel tárgyát képező sokaság egyedeit megfigyelési egységeknek nevezzük (azon egyedet tehát, akire/amikre vonatkozóan adatokat/információkat gyűjtünk). Ezek az egyedek nem feltétlenül azonosak az adatszolgáltató, az ún. **számbavételi egységekkel**. Például

dául állatszámmlálás esetén a megfigyelési egységek az egyes állatok, a számbavételi egységek pedig az egyes gazdálkodók, vállalkozók.

A kérdőíveket **önszámlálás** esetén maga az adatszolgáltató tölti ki, **kikérdéses eljárásnál** a számlálóbiztosok jegyzik fel a válaszokat.

A statisztikai adatfelvételek egyik kulcskérdése a kérdőívek helyes megszerkesztése, ami a módszertani ismeretek mellett az adott terület alapos szakmai ismeretét is igényli. A feltett kérdéseknek egyértelműeknek, közérthetőeknek kell lenniük, és igazodniuk kell a vizsgálat céljaihoz. A nem eléggé körültekintően megfogalmazott kérdések ugyanis a valóságtól eltérő irányba terelhetik a válaszadást.

Az előzőekből következik, hogy minden adatfelvétel bizonyos hibalehetőséget rejt magában. Hibát eredményezhet a pontatlan kérdéseket, fogalmakat tartalmazó kérdőív, az adatszolgáltató valóságostól eltérő válaszai, a szervezési-végrehajtási hibák.

Az eddig leírtakból látható, hogy **a statisztikai adatok** általában csak **korlátozottan pontosak** lehetnek. Egyrészt a már említett adatfelvételi hibák szinte elkerülhetetlen fellépése miatt, másrészt az adatfeldolgozás és adatközlés során előforduló esetleges tévedések miatt. Ezért a valóságos (pontos) adat és a hibákkal torzított mért adat egymástól eltér.

A valóságos adat (A) és a mért adat (\hat{A}) különbségét a statisztikai adat **abszolút hibájának** nevezzük, és a -val jelöljük:

$$a = |A - \hat{A}|.$$

A gyakorlatban az abszolút hibát nem tudjuk meghatározni, mivel a valóságos adat (A) nem ismert. Ezért becslést adunk arra a számértékre, amelynél az abszolút hiba biztosan nem nagyobb.

Az adott becslést a közelítő érték (mért adat) **abszolút hibakorlátjának** (\hat{a}) nevezzük. Így minden statisztikai adat megadható az $\hat{A} \pm \hat{a}$ módon. Ez a megadási mód arra utal, hogy a valóságos adat (A) valahol az $\hat{A} - \hat{a}$ és $\hat{A} + \hat{a}$ határok között helyezkedik el.

A statisztikai adatok $\hat{A} \pm \hat{a}$ módon történő megadása helyett a gyakorlatban igen elterjedt megoldás az is, hogy a statisztikai adatokat bizonyos nagyságrendre kerekítve adjuk meg, azaz a statisztikai adatban számszerűen csak az ún. **szignifikáns számjegyek** jelennek meg. Szignifikáns számjegyeknek nevezzük azokat a számjegyeket, amelyekben még feltétlenül megbízunk, amelyeket még pontosnak fogadunk el.

Ha a legutolsó kiírt számjegy helyi értéke 10^{sz} , akkor az abszolút hibakorlát:

$$\hat{a} = \frac{10^{sz}}{2}.$$

Egy magyarországi kft. formájában bejegyzett kisvállalkozás 2012. évi eredménykimutatásából kiolvasható éves nettó árbevétele 12 248 ezer Ft. A közölt statisztikai adatban a legutolsónak (számszerűen) kiírt szignifikáns számjegy helyi értéke 10^3 . (Ez a közlési mód azt sugallja, hogy az utolsó három – száz, tízes, egyes helyi értékű – számjegy nem megbízható [nem szignifikáns], ezért számszerűen nem írjuk ki.)

A statisztikai adat (nettó árbevétel) abszolút hibakorlátja:

$$\hat{a} = \frac{1}{2} \cdot 10^3 = 500 \text{ Ft.}$$

Gyakran célszerűbb, kifejezőbb az elkövetett hibát (vagy hibakorlátot) a valószínűségi (vagy mért) adathoz viszonyítani.

Az abszolút hiba (a) és a valószínűségi adat (A) hányadosát **relatív hibának**,

$$\alpha = \frac{a}{A},$$

az abszolút hibakorlát (\hat{a}) és a mért adat (\hat{A}) hányadosát a statisztikai adat **relatív hibakorlátjának** nevezzük.

$$\hat{\alpha} = \frac{\hat{a}}{\hat{A}}.$$

A relatív hibát és a relatív hibakorlátot általában százalékos formában szokták megadni.

Példánkban a nettó árbevétel relatív hibakorlátja:

$$\hat{\alpha} = \frac{500}{12\,248\,000} = 0,00004082 \rightarrow 0,004\%.$$

Megállapíthatjuk, hogy a fenti példában a statisztikai adat hibája rendkívül kicsi.

A korlátozott pontosságú (pontatlan) statisztikai adatokkal végzett minden számítási művelet eredménye ugyancsak korlátozottan pontos (pontatlan) lesz. Ezért mind az adatok kezelésénél, mind a belőlük levont következtetéseknél figyelembe kell vennünk az adatok korlátozott pontosságát. Így elkerülhető például, hogy nem szignifikáns eltérések alapján rangsorolást végezzünk vagy nem valós különbségeket magyarázzunk.

1.4. Statisztikai csoportosítás és összehasonlítás

Csoportosítás

A statisztikai megfigyelés eredményeként nagyszámú adathoz jutunk, amelyek a vizsgált sokaságról különböző ismérvek alapján nyújtanak széles körű információt, számszerű ismereteket. Ahhoz, hogy a sokaságot, annak összetételét megismerhessük, a sokaságot a különböző ismérvek szerint osztályoznunk, csoportosítanunk kell.

A csoportosítás a sokaság felosztása a sokaság egységeit jellemző megkülönböztető ismérv szerint.

A csoportosításnál ügyelni kell arra, hogy olyan sokaságrészeket, ún. **osztályokat** alakítsunk ki, hogy azok **átfedésmentesek** és **teljesek** legyenek. E két követelmény együtt azt jelenti, hogy a sokaság minden egysége egyértelműen besorolható legyen valamelyik – de csak egy – kialakított osztályba.

Ha a csoportképző ismérv változatainak száma kevés és egymástól pontosan elkülöníthető (pl. ha az aktív keresőket nemek vagy megyék szerint csoportosítjuk), az osztályok képzése nem okoz gondot. Ilyen esetben általában *egy* ismérvváltozat képez egy osztályt. Ha az ismérvváltozatok száma nagy, az osztályok képzése már nem egyértelmű, és a módszertani ismereteken túl szakmai ismereteket is igényel. (Pl. az aktív keresők foglalkozás, kereset szerinti, vagy a vállalkozások tevékenységtípus szerinti csoportosításánál.) Ha a vállalkozások esetén minden tevékenységtípust felsorolnánk, egy hosszú „listát” kapnánk, ami nehezen áttekinthető. Ilyen esetben szükség lehet arra, hogy az adott ismérv *egynél több* változata képezzen egy osztályt.

A gyakorlatban ilyen csoportosításoknál általában az ún. **nómenklatúrákat** – szabványnak tekinthető, rendszeres felhasználásra kerülő osztályozási rendszereket – alkalmazzák.

Az egy ismérv szerinti osztályozás eredménye egy **csoportosító sor**, melynek általános sémája (1.2. táblázat):

1.2. TÁBLÁZAT
Csoportosító sor általános sémája

Osztály	Egységek száma
C_1	f_1
C_2	f_2
\vdots	\vdots
C_i	f_i
\vdots	\vdots
C_k	f_k
Összesen	N

Ahol: C_i : a csoportképző ismérv alapján képzett i -edik osztály azonosítója,
 f_i : a sokaság C_i osztályba sorolt egységeinek száma, ún. **gyakorisága**,
 k : a kialakított osztályok száma,
 N : a sokaság egységeinek száma.

A sémából is látható, hogy a gyakoriságok (f_i) összege egyenlő a sokaság elemszámával (N).

Tehát

$$f_1 + f_2 + \dots + f_k = \sum_{i=1}^k f_i = N.$$

A csoportképző ismérv fajtájától függően a csoportosító sorok lehetnek: **minőségi, mennyiségi, területi** és **idősorok**. Az idősorok képzésével a fejezet végén, a mennyiségi sorok képzésével, jellemzőivel a tankönyv 2. fejezetében részletesen foglalkozunk, ezért itt csak a minőségi és területi sort szemléltetjük egy-egy példával.

Minőségi sor (1.3. táblázat):

1.3. TÁBLÁZAT

A magyarországi vendéglátóhelyek bevételei a vendéglátóhelyek típusa szerint 2011-ben

Vendéglátóhelyek típusa	Bevétel, millió Ft
Kereskedelmi vendéglátóhelyek	599 776
Munkahelyi vendéglátóhelyek	98 208
Összesen	697 984

Forrás: Magyar Statisztikai Évkönyv, 2011, 354. o.

Területi sor (1.4. táblázat):

1.4. TÁBLÁZAT

A Magyarországra bevándorló külföldi állampolgárok száma földrészek szerint 2011-ben

Földrész	Állampolgárok száma, fő
Európa	16 518
Ázsia	3 824
Amerika	1 524
Afrika	589
Ausztrália és Óceánia	59
Összesen	22 514

Forrás: Magyar Statisztikai Évkönyv, 2011, 32. o.

Az egy ismerv szerinti csoportosítás a sokaságról kevés információt nyújt, ezért gyakran alkalmazzuk az ún. **kombinatív csoportosítást**. Ennek lényege, hogy az egyik ismerv szerint képzett osztályokon belül egy másik ismerv szerint is csoportosítunk. Például a lakásépítéseket csoportosítjuk építető és építési cél szerint Magyarországon 2011-ben.

Természetes személy 8007				Helyi és központi költségvetés 137				Jogi személyiségű 4392				Egyéb 119			
Saját használat	Bérbé- adás	Értéke- sítés	Egyéb	Saját használat	Bérbé- adás	Értéke- sítés	Egyéb	Saját használat	Bérbé- adás	Értéke- sítés	Egyéb	Saját használat	Bérbé- adás	Értéke- sítés	Egyéb
7538	47	410	12	-	111	11	15	39	30	4286	37	4	1	105	9

Forrás: KSH Tájékoztatási adatbázis, <http://statinfo.ksh.hu/Statinfo/haViewer.jsp>, *Technikai azonosító: RB1A01_W*
Épített lakások. Az országban épített összes lakás.

A kombinatív csoportosítással kapott adatokat táblázatba is rendezhetjük, **statisztikai táblát** készíthetünk. Ennek részleteiről a 3. fejezetben lesz szó.

Összehasonlítás

A mindennapi életben és a statisztikai elemző munkában is gyakran találkozunk az összehasonlítás mozzanatával.

Az **összehasonlítás két vagy több statisztikai adat egymáshoz való viszonyítása**.

Az összehasonlítás az adatok egyszerű összevetésén túl általában különbség és hányados képzésével történik. Például a tartós lekötésű betétek devizaszámla-kamata az OTP Nyrt. Banknál 2011. június 23-án 3,437%, július 23-án 3,375% volt. Az adatok pusztá összevetése alapján azt tudjuk megállapítani, hogy a devizaszámla-kamat csökkent. Ha a változás nagyságára is kíváncsiak vagyunk, akkor a két időpont devizaszámla-kamatának a különbségét vagy hányadosát számítjuk ki.

A két szám különbsége: $3,375 - 3,437 = -0,062$.

A két szám hányadosa: $\frac{3,375}{3,437} = 0,9819 \rightarrow 98,2\%$.

A devizaszámla-kamat 0,062 százalékponttal, illetve 1,8 százalékkal csökkent.

Két, százalékban (ezrelékben) kifejezett adat (mutatószám) különbségének mértékegységét **százalékpontnak (ezrelékpontnak)** szokás nevezni. (A kamat-változás mértékét a gyakorlatban általában százalékpontban adják meg.)

Az összehasonlítandó adatokat is statisztikai sorba rendezhetjük. Az így képzett sorokat **összehasonlító soroknak** nevezzük, melyeket – a csoportosító sorokhoz hasonlóan – az ismérvek fajtája szerint is megkülönböztethetünk.