

GRAFIKUS ÁBRÁZOLÁS A STATISZTIKÁBAN

HUNYADI LÁSZLÓ

A cikk a grafikus ábrázolás néhány kérdését tekinti át. Kiinduló pontja az, hogy bár a grafikus ábrázolás a statisztika egyszerűbb eszközei közé tartozik és ennél fogva igen népszerű, sok a pontatlan, a félreérthető, sőt egyenesen hibás ábrázolás. A ma rendelkezésre álló technika, az elterjedt szoftverek igen sok ábrázolási lehetőséget kínálnak ugyan a felhasználónak, de egyrészt az ezek közti választás nem mindig egyszerű, másrészt még helyes választás esetén is gyakoriak a hibás alkalmazások. A tanulmány példák nagy tömegén keresztül próbálja bemutatni a helyes és a hibás gyakorlatot.

A legegyszerűbb ábrázolási lehetőségeken túlmenően a cikk felhívja a figyelmet arra, hogy egy sor olyan ábra áll az elemzők rendelkezésére, melyek bonyolultabb, ám gyakran előforduló statisztikai feladatok esetén nyújtanak plasztikus bemutatási lehetőséget. Az itt látható alkalmazások azt is sugallják, hogy egy kis találékonysággal, intuícióval gyakorlatilag minden jelenségre szemléletes célábra készíthető.

TÁRGYSZÓ: Grafikus ábrázolás. Statisztikai programcsomagok.

A statisztikai elemzések és közlések egyik fontos, bár sokat vitatott eszköze a grafikus ábrázolás. Sokat vitatott, hiszen szép számmal vannak hívei és ellenzői egyaránt. Akik az ábrákat szeretik, túl azon a közhelyen miszerint „vizuális típus vagyok” – feltehetően felismerik akár öntudatlanul is azt –, hogy a megfelelően megszerkesztett ábrák a statisztikai munka különböző fázisainak igen tömör és robusztus eszközei. Akik ellenzik az ábrákat, többnyire az azok által hordozott puha információkat kifogásolják, és arra utalnak, hogy az ábrázolás igen sok visszaélési lehetőséget rejt magában. Ha a statisztika segítségével mindent és mindennek az ellenkezőjét is be lehet bizonyítani – ahogy mondják –, akkor fokozottan áll ez az ábrázolásra.

Valószínű, hogy mint sok mindenben, ebben a kérdésben is valahol a két szélsőséges vélemény között kell keresni a helyes utat: az ábrázolás a statisztika fontos és nélkülözhetetlen eszköze, és mint ilyen, szerves részét kell képeznie az elemzéseknek. Ugyanakkor az ábráknak nem szabad túltengeniük, és ami a legfontosabb, mindent el kell követni annak érdekében, hogy az ábrázolás korrekt és félreérthetetlen információt adjon, ne tegyen lehetővé tudatlanságból vagy sanda szándékból adódó hibás interpretációt.

A grafikus ábrázolás igen hosszú múltra tekinthet vissza a statisztikában, hiszen már jó 200 éve (lényegileg amióta a mai értelemben vett statisztika létezik) rendszeresen használják. A grafikus ábrázolással sokat foglalkozott és foglalkozik a szakirodalom.

Részletes szakirodalmi hivatkozás helyett talán elegendő utalni a statisztikai tudomány enciklopédiájára (*Encyclopedia of Statistical Sciences*; 1993), amely kiváló történeti áttekintést ad, és jól mutatja be az aktuális helyzetet.¹ Annak, hogy mindezek ellenére ezt a témát újból elő kell venni, több oka van.

– A felgyorsult és információkkal túlsordulásig telt világban minden eddiginél nagyobb szükség van az információk tömör, gyorsan áttekinthető, ugyanakkor korrekten megjelenítésére. Az emberek szeretnek gyorsan olvasni, néhez munka nélkül hozzájutni bizonyos információkhoz. Ezért a grafikus módszerek az utóbbi években felértékelődtek, így minden alkalmat meg kell ragadnunk arra, hogy előmozdítsuk a grafikus ábrázolás helyes gyakorlatát.

– A számítógépek és hálózatok rohamos terjedése korábban elképzelhetetlen lehetőségeket tárt fel a statisztikai ábrázolás területén. Ezek teljes körű áttekintése nem tárgya ugyan a jelen írásnak, de feltétlenül fel kell hívni a figyelmet az ebben rejlő lehetőségekre.

– Mind a számítógépes, mind a hagyományos ábrázolás terén a hazai társadalom- és gazdaságstatisztikai gyakorlat meglehetősen beszűkült. Sajnos ugyanez mondható el az egyéb, nem grafikus eszközök használatáról is, hiszen az elemzések igen nagy része megragad a legegyszerűbb mutatószámok (átlag, hányados, egyszerű viszonyszámok) kiszámításánál és értékelésénél. Hasonló a helyzet az ábrázolásnál is: az elemzések döntő hányada csak a legegyszerűbb ábratípusokat használja, holott hosszadalmas magyarázatok, számítások, táblázatok helyett gyakran jóval többet mondana egy-két speciális ábra bemutatása. Ezért – úgy gondoltuk – fel kell hívni a figyelmet arra, hogy a grafikus ábrázolás lehetőségei jóval szélesebbek annál, mint amit ma tömegesen használunk.

– Foglalkozni kell a grafikus ábrázolással azért is, mert éppen a gépi eszközök, mindenki számára hozzáférhető szoftverek szaporodása folytán a felhasználók nagymértékben elbizonytalanodtak. Az egyes programcsomagok által felkínált ábratípusok nagy száma igen gyakran olyan alkalmazásokra csábítja a felhasználókat, amelyek nem felelnek meg az adott feladatnak, jó esetben csak szokatlanok, nehezen áttekinthetők, rosszabb esetben hibás, félrevezető következtetésekhez vezetnek.

Mindezek miatt célszerűnek látszik a grafikus ábrázolás lehetőségeinek kritikai áttekintése. Ahhoz, hogy ezt a széles területet vagy annak legalább lényegesebb részeit tervszerűen be tudjuk járni, valamilyen rendező elvet kell találnunk. Az említett enciklopédia nagyon részletes csoportosítást közöl, mely a grafikus ábrázolás eszközeit a statisztikai munka különböző fázisaiban betöltött szerepe szerint csoportosítja. Ennek lényege egy ötfokozatú osztályozás, amely az ábrázolás céljait a következőkben jelöli meg:

- exploratív célú (a jelenségek kapcsolatait, okait stb. tisztázó, kereső) alkalmazás;
- leíró célú alkalmazás;
- alkalmazás döntés-előkészítéshez;
- közlés, a felhasználók különféle rétegeinek tájékoztatása;
- a statisztikai munka belső eszközeként való alkalmazás.

Ez a csoportosítás tovább finomítható, ám mivel célunk nem a teljes körű áttekintés, inkább egyszerűsítjük a felosztást, és a későbbiekben csak *munkaábráról*, illetve *közlési ábráról* beszélünk. Az előbbi kategóriába tartoznak mindazok az ábrák, amelyek a statisztikus munkáját segítik, de nem feltétlenül érdekesek a felhasználó és kiváltképp nem a szélesebb olvasóközönség számára. A közlési ábrák ezzel szemben a legszélesebb felhasználói kör számára adnak vizuális információt. Megjegyezzük, hogy a két kategória közt a határvonal nem éles, gyakran egy ábra akár ide, akár oda is tartozhat. Példaként említhetünk egy grafikus tesztet, amely a statisztikus munkája közben tájékoztathatja, és segíthet megtalálni a helyes modellt, ugyanakkor lehet az elemzés olyan végeredménye is, amely szélesebb ér-

¹ Jóllehet ez az összeállítás már közel 20 éves, az azóta eltelt évek inkább csak a speciális számítógépes ábrák terén hoztak lényeges fejlődést, ezekkel pedig itt csak érintőlegesen foglalkozunk. Ezért a téma tanulmányozásához jó kiindulási alap az említett áttekintés.

deklódásra tarthat számot. A későbbiekben először a szélesebb érdeklődésre számot tartó ábrákkal, majd ezt követően a speciális célábrákkal foglalkozunk.

A téma feldolgozása során figyelmünket a helyes, illetve a hibás alkalmazások felé fordítjuk, és lényegileg három elkövethető hibát különböztetünk meg:

- a mondanivaló szempontjából nem megfelelő ábratípus kiválasztása;
- a tárgynak megfelelő ábra helytelen használata;
- kimaradó lehetőségek, melyek jobbára abból adódnak, hogy a felhasználók nincsenek tisztában azzal, hogy milyen ábrák állnak rendelkezésükre.

A tanulmány további részeit lényegileg e három hibatípus köré csoportosítottuk. A bevezetőt követő fejezetben bemutatjuk az ábrázolás általunk legfontosabbnak tartott alapelveit, majd a jelenlegi gyakorlatban leginkább elterjedt ábratípusok közti választással foglalkozunk. Eközben rámutatunk a hibás, félrevezető vagy legalábbis nem szerencsés alkalmazásokra, és végezetül csokorba szedjük azokat a fontosabb ábratípusokat, amelyek kívül esnek a felhasználók legnagyobb részének jelenlegi gyakorlatán, jóllehet megismerésük és alkalmazásuk gazdagíthatná a hagyományos elemző munkát.

Szólni kell a tanulmány módszeréről is. Csábító lehetőség ilyen esetekben az, hogy a közelmúlt jó és rossz példáit idézve mutatunk rá a helyesnek vélt gyakorlatra. Tekintve azonban, hogy a legkevésbé sem célunk egyes szerzők vagy csoportok kipellengérezése, inkább azt a módszert választottuk, hogy a rossz, elrettentő példákat külön úgy alkottuk meg, hogy a helytelen alkalmazáson túlmenően azok semmiben se emlékeztessenek élő vagy elhunyt szerzők munkáira. A helyes, követendő gyakorlatot is gyakran ismertetjük saját, e célra készült egyszerű példákon, de ez esetben nem zárkozunk el egy-egy jól sikerült alkalmazás bemutatásától. Feltétlen említést érdemel az, hogy az adatforrás megjelölése általában a kiinduló adatokra, nem pedig az esetlegesen hibás ábrázolásra utal. A két esetet megkülönböztető az *adatforrás* és a *forrás* tehát két különböző fogalmat takar.

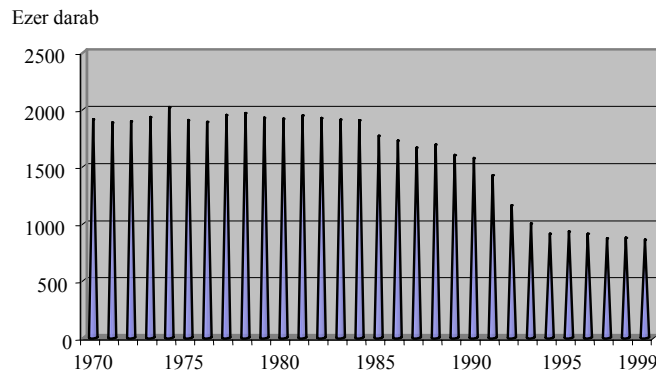
A GRAFIKUS ÁBRÁZOLÁS ALAPELVEI

Mielőtt az egyes jelenségek bemutatására szolgáló ábratípusokat részletesen is bemutatnánk, röviden szólnunk kell azokról az alapelvekről, melyeket minden ábrázolásnál szem előtt kell tartanunk. A szakirodalomban több ilyen is megtalálható, természetesen az itt felsorolandók is bővíthetők, szűkíthetők, vitathatók. Felfogásunk szerint tehát a legfontosabb, általános normák, követelmények a következők.

a) Az ábra legyen *áttekinthető*, azt és csak azt mutassa, amire szolgál. Ez az elv meglehetősen elcsépeltnak tűnhet, de valójában nem az. Gyakori, hogy az alkalmazók az ábrával mást, többet, szebbet akarnak bemutatni, mint amit valójában kellene, lehetne. Kiváltképp a számítógépes ábrák csábítanak szép, de értelmetlen formákra, feleslegesen kevert színekre és árnyalatokra, a tárgy szempontjából lényegtelen, sokszor zavaró formai megoldásokra. Ügyelni kell arra is, hogy a színes gépi ábrák fekete-fehér nyomtatásban olyan árnyalatokat kaphatnak, amelyek az egyébként jól megkülönböztethető színeket összemossák.

Az 1. ábra, amely a sarvasmarha-állomány alakulásának idősorát tartalmazza, nem tesz eleget az áttekinthetőség követelményének, hiszen az alkalmazott jel (gúla) elvonja a figyelmet a lényegről, ráadásul az idomok változó vastagsága is zavaró, emellett a háttérszín a fekete-fehér ábrán rontja a kiemelést.

1. ábra. Magyarország szarvasmarha-állománya, 1970–1999

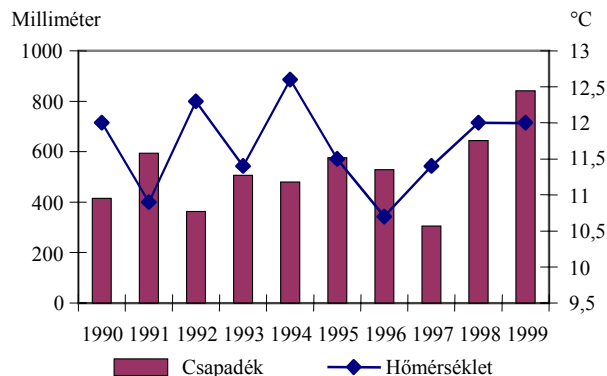


Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

b) Az ábra legyen *célorientált* és *homogén*, lehetőleg egy jelenséget ábrázoljon. Az utóbbi időkben nagy népszerűsége tettek szert az ún. kétszögű ábrák. Ezek lényege az, hogy egy koordináta-rendszerben két, esetleg több jelenséget akarnak egyszerre bemutatni. Két eltérő skálájú tengely, az egy ábrában együtt megjelenő eltérő ábratípusok (például vonaldiagram és oszlopdiaagram) felfogásunk szerint sokkal inkább zavarják, mintsem segítik a gyors megértést. Így van ez még akkor is, ha a jelenségek összetartoznak, és az ábra éppen ezt a kapcsolatot hivatott bemutatni. A kettős skálázás, ami nyilvánvalóan a különböző mértékegységekből, illetve eltérő nagyságrendekből adódóan szükséges lehet, nehezen áttekinthetővé, nehezen értelmezhetővé teszi az ábrát, és így éppen legfontosabb előnyétől fosztja meg ezt az eszközt.

A 2. ábrán két különböző mértékegységben mért, de egymással némileg összetartozó jelenséget zsúfoltunk össze: a vonallal összekötött pontok az elmúlt tíz év átlaghőmérsékletét, az oszlopok az éves csapadékmennyiséget mutatják. Az ábra túl sok információt hordoz, ezért áttekinthetetlen. A két tengely eltérő skálája összezavarja a képet, nem segíti a jelenség jobb, gyorsabb megértését, ezért alkalmazását nem javasoljuk.

2. ábra. Az éves átlaghőmérséklet és az évi csapadék mennyisége Budapesten

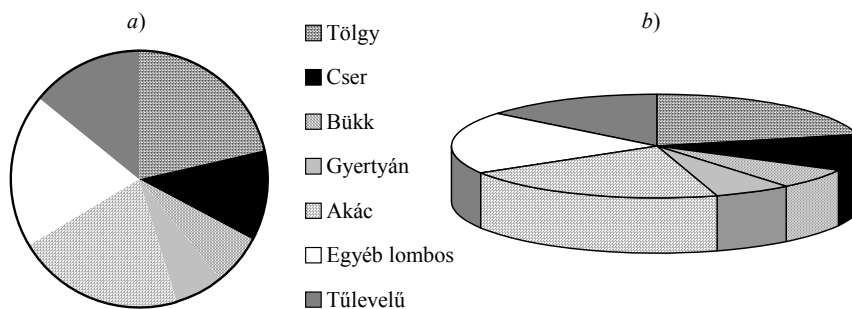


Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

c) Az ábra legyen a lehető *legegyszerűbb*. Napjainkban a számítógépes ábrák kapcsán elterjedtek a háromdimenziós (3D) ábrák. Egyszerű paranccsal kérhető 3D hisztogram vagy kördiagram, és ezek olykor szebben mutatnak, mint kétdimenziós párjaik. Ha azonban a harmadik dimenziónak semmiféle funkciója nincs, ez az ábrázolás csak bonyolítja az áttekintést, elvonja a figyelmet, akadályozza a jelenség gyors felismerését, ezért nemcsak hogy felesleges, de kifejezetten káros is.

Az 3. ábra a magyarországi erdőterület 1999. évi fajtankénti megoszlását mutatja. Az a) kördiagramon jól látható a fő fajták (tölgy, akác, egyéb lombos) túlsúlya, és egymáshoz viszonyított arányuk is világosan leolvasható az ábráról. Ezzel szemben a b) kördiagram, amellyel ugyanezt a jelenséget egy 3D ábrán interpretáljuk, egyrészt feleslegesen tartalmazza a harmadik dimenziót (a korong magasságát), és ezzel megosztja a figyelmet, másrészt kifejezetten zavaró az, hogy nem szimmetrikus helyzeténél fogva az akác arányát látszólag nagyobbak, az egyéb lombos és a cser arányát kisebbnek tünteti fel a valóságnál. Emellett a térhatás kiemelésére hivatott árnyékolás alkalmat ad az egyes fajták jelölésének összekeverésére.

3. ábra. A magyarországi erdőállomány fajta szerinti megoszlása 1999-ben



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

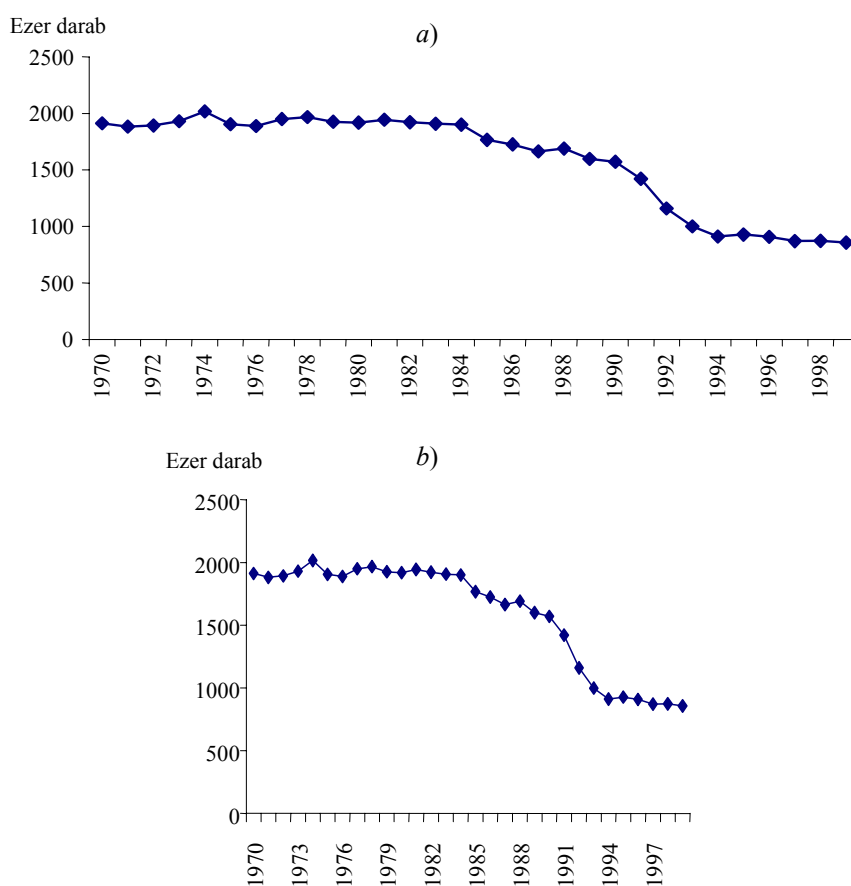
d) Az ábrának, akárcsak bármilyen elemzésnek *rekonstruálhatónak* kell lennie, ami azt jelenti, hogy az adatok forrására és a fontos módszertani elemekre lehetőleg az ábrán, de legalább a hozzá kapcsolódó szövegben pontos utalásnak kell lennie. Ide tartozik az is, hogy az ábrát pontos és egyértelmű jelmagyarázattal kell ellátni. Több szerző kifejezetten elvárja a jelmagyarázattól, hogy az a szövegtől a lehető legnagyobb mértékben független, azaz az ábra szöveggörnyezetéből kiemelve is informatív legyen. Ez – úgy véljük – vitatható, hiszen bonyolultabb esetekben olyan sok szöveget jelentene, hogy éppen a sok magyarázat tenné értelmetlennek az ábrát. Mindazonáltal a pontos cím, a tengelyek mértékegységének és skálájának egyértelmű megjelölése, a forrás megadása, valamint a tömör, de a szöveges rész alapján érthető jelmagyarázat minimális követelménynek tekinthető.

e) Az ábrát úgy kell méretezni (skálázni), hogy az *optikailag semleges*, a lehető legnagyobb mértékben konform legyen az elemzés mondanivalójával, ugyanakkor ne legyen különféle grafikus trükkökkel befolyásolni. A grafikus ábrázolásban az egyik legtöbb veszélyt rejtő helyzet, amikor a derékszögű koordináta-rendszer két tengelye más mértékegységben adott. Ekkor ugyanis a skálázás az elemzőre van bízva, és a tengelyek skálabeosztását változtatva minőségileg más-más következtetés vonható le ugyanabból az ábrából. Ilyen esetekben célszerű több (legalább 2) skálázással bemutatni az ábrát, elkerülendő a manipulációnak még a gyanúját is. Amennyiben a jó elrendezés megköveteli

azt, hogy a tengelyeket megszakítsuk, ezt világosan és egyértelműen jelölni kell az ábrán, hiszen ellenkező esetben az arányok torzulhatnak, és a felületes szemlélőt félrevezetheti az ábra. (Lásd például a 21. ábrát.) Néha a nagy szélsőségek kiegyenlítése érdekében valamely (esetleg mindkét) tengelyen logaritmikus skálázást alkalmazunk. Ez munkaábrák esetében olykor hasznos lehet, de nehezen értelmezhető, és annyira megváltoztatja a képet, hogy alkalmazását közlési ábrák esetén nem javasoljuk.

A 4. ábra a korábban már ismertetett szarvasmarha-állomány idősrát mutatja be, ezúttal pontdiagrammal. Ez az ábrázolási mód megfelel a jelenség természetének, ám a két ábrán más a skálabeosztás, ami minőségileg eltérő következtetésre vezethet, ha valaki csak az a) változat és a b) változat alapján akarja értékelni a jelenséget. Az a) egy stabil, tartósan, de enyhén csökkenő állományt mutat, míg a b) ábra első ránézésre erőteljes, zuhanásszerű változást sejtet.

4. ábra. Magyarország szarvasmarha-állománya, 1970–1999



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

f) Bizonyos ábratípust lehetőleg csak egyfajta jelenség bemutatására használjunk; ez esetben ugyanis már maga az ábratípus is ad a felhasználónak lényeges, hasznosítható és

kényelmesen hozzáférhető információt. Az ábratípus helyes megválasztása nem könnyű, sőt nem is mindig egyértelmű feladat, de alapvető fontosságú, hiszen a típus hibás megválasztása eleve lehetetlenné teszi a helyes ábrázolást. Ezért a következőkben megkíséreljük összefoglalni azokat a tudnivalókat és konvenciókat, amelyek alapján egyszerűbb esetekben nagy biztonsággal ki lehet választani a vizsgálandó jelenséghez a megfelelő ábratípust.

AZ ALKALMAZOTT FŐ ÁBRATÍPUSOK

A megfelelő ábratípus kiválasztásánál csak azokat az egyszerűbb eseteket vesszük sorra, amelyek szinte minden elemzésben előfordulhatnak. A bonyolultabb esetekben bizonyos értelemben könnyebb a választás (hiszen az ábrák is speciálisak), és az alkalmazó statisztikus is általában képzetesebb, így ritkábban okoz számára nehézséget a helyes ábrázolási mód megválasztása.

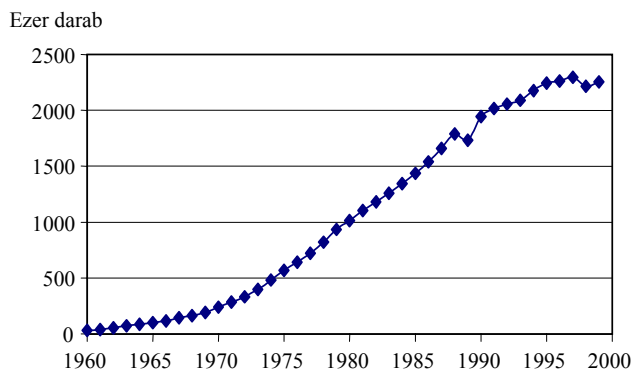
A legegyszerűbb ábrákat statisztikai sorok és viszonyszámok bemutatására használjuk. Ezen egyszerű mutatószámok rendszerező áttekintése megtalálható az alapfokú statisztikai könyvekben (például *Hunyadi et al.*; 2000), ezért ezzel itt nem foglalkozunk. Ehelyett csak a leggyakoribb esetek bemutatására szorítkozunk.

Talán a leggyakoribb az *idősorok* ábrázolása, amelynél bevett szokás az, hogy az idősort meghatározó időbeli ismérv értékét derékszögű koordináta-rendszer vízszintes tengelyén mérjük fel, és az idő előrehaladtát balról jobbra ábrázoljuk.

Állapotidősorok esetén az időbeli ismérv értékei egy-egy időponthoz tartoznak, ezért célszerű ábrázolásuk egy-egy pont. Az állapotidősor javasolt ábrája tehát a pontdiagram, amelynél az egyes pontokat egyenesekkel össze is lehet kötni. Ez egyrészt utalhat arra, hogy a jelenség a megfigyelt időpontok közt is létezik (és a szakaszokkal egyszerű lineáris interpolációs értéket rendelünk minden időponthoz), de szolgálhat egyszerűen arra is, hogy az idősorban meglévő tendencia jobban láthatóvá váljék.

Az 5. ábra a magyarországi személygépkocsi-állomány 1960 és 1999 közötti alakulását mutatja. Az egyes időpontokhoz tartozó állományok adatait pontokkal ábrázoljuk. Az adatok az év végére vonatkoznak, de ez a kis lépték miatt nem látszik. A pontokat egyenesekkel kötöttük össze, így még jobban kirajzolódik az alap tendencia: az egyre gyorsuló állománynövekedés az 1990-es évek közepétől a telítettség jeleit kezdi mutatni.

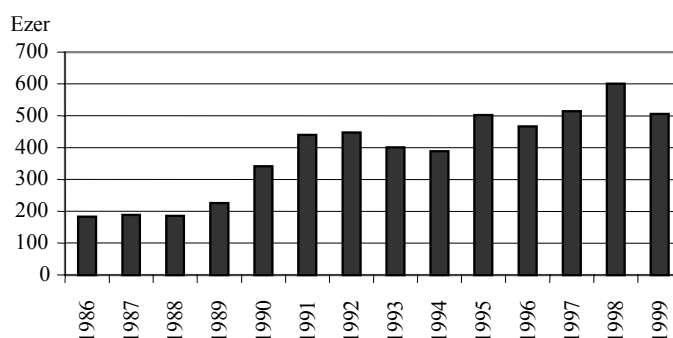
5. ábra. A magyarországi személygépkocsi-állomány alakulása, 1960–1999
(év végi állomány)



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Tartamidősorok esetén a vízszintes tengelyen elvben intervallumok szerepelnek, a jelenséget pedig célszerű ezen intervallumok fölé rajzolt téglalapokkal (oszlopokkal) bemutatni. Ha (és ez a jellemző) a megfigyelési intervallumok egyenlő hosszúságúak, akkor a téglalapok magassága arányos a jelenség mindenkori jellemzőjének értékével. Attól függően, hogy a megfigyelési időintervallumok megszakítás nélkül követik-e egymást, vagy van köztük kihagyás, az oszlopok közvetlenül egymáshoz simulnak, illetve hézag van köztük. Mivel az esetek nagy részében folyamatos idősorokról van szó, hézag nélküli oszlopokkal kellene a jelenséget ábrázolni. Az ilyen ábrázolás azonban formailag megegyezik a hisztogramos ábrázolással. Tekintettel arra, hogy – mint azt később látni fogjuk – a hisztogramnak kitüntetett jelentőséget tulajdonítunk az ábrázolás során, az „egy jelenség – egy ábratípus” elv alkalmazása azt sugallja, hogy ilyen esetekben hézaggal illeszszük az oszlopokat, ezzel elkerülve a hisztogrammal való összekeverést. A 6. ábra egy tartamidősor ábrázolását mutatja be.

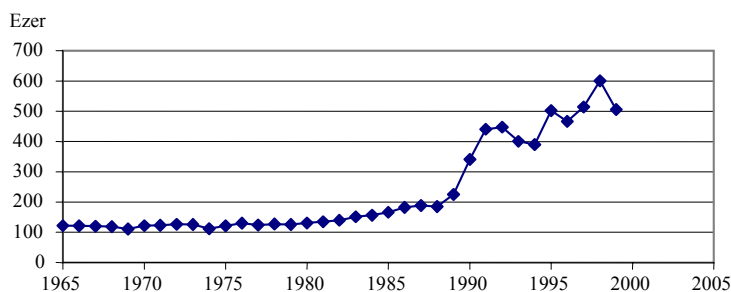
6. ábra. A bűncselekmények számának alakulása Magyarországon, 1986–1999



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Megjegyezzük, hogy a gyakorlatban tartamidősorok esetén is gyakori a pontdiagrammal történő ábrázolás. Ez – bár elvileg helytelen – nem nagyon zavaró, alkalmazása elterjedtsége miatt aligha zárható ki, ezért elfogadható. Ebben az esetben azonban igen fontos, hogy a jelenség alakulását leíró változó értékei ne az időszak elejéhez vagy végéhez, hanem a megfelelő intervallum közepéhez tartozzanak.

7. ábra. A bűncselekmények számának alakulása Magyarországon, 1965–1999

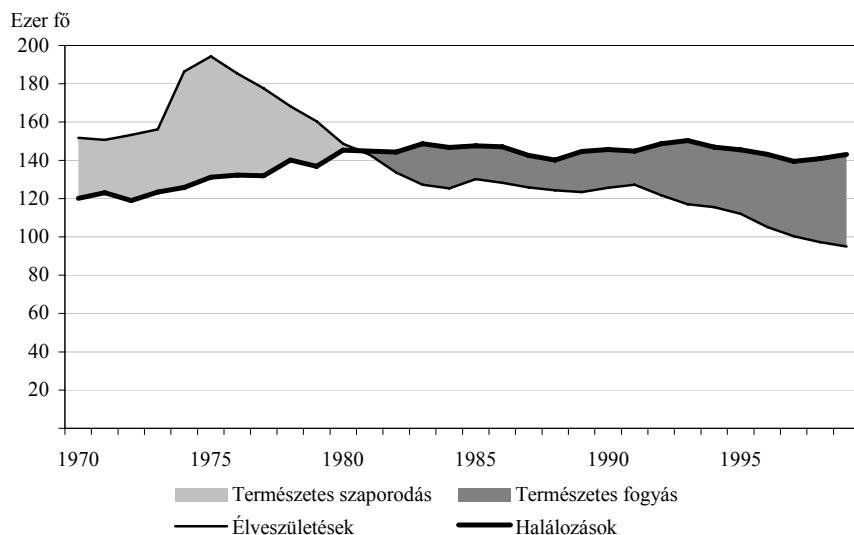


Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Idősorok esetén kiváltképp fontos figyelni a korábban említett skálázási problémára, hiszen itt biztos, hogy a két tengely eleve más mértékegységben van kifejezve. Az idősorok ábrázolásánál arra is ügyelni kell, hogy ha az idősort valamilyen módon modellezzük (például trenddel), szűrjük, szezonálisan kiigazítjuk, ezt általában folytonos és eltérő jelzéssel ellátott vonallal kell az ábrában jelölni.

Két vagy több idősor esetén egyes esetekben szemléletesen lehet ábrázolni kapcsolatokat, mint ahogy azt a 8. ábra is mutatja. Az ábrán a két idősor (az élveszületések és a halálozások száma) egyenlegeként előálló árnyékolt terület igen szemléletesen mutatja a népesség szaporodásának, illetve fogyásának fő összetevőit.

8. ábra. Élveszületések, halálozások, szaporodás és fogyás, 1970–1999



Forrás: Magyarország, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Területi sorok esetén az ábrázolás többnyire térkép segítségével történik. A jelenséget – legyen az bármilyen területi összehasonlítás – a térképen, vagy az egyes területek megfelelő színezésével, vagy az egyes területekre (megyék) illesztett oszlopokkal vagy egyes pontok kirajzolásával ábrázoljuk. Természetesen más, térképen alapuló megoldások is elképzelhetők.

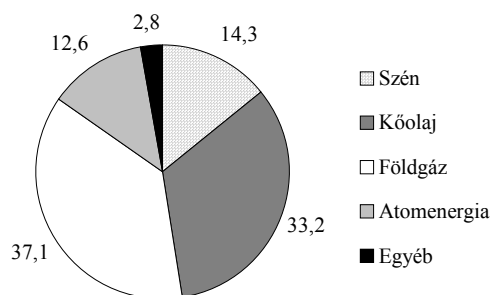
A térképes ábrázolással kapcsolatosan két, nem túl jelentős hibaforrásra hívjuk fel a figyelmet. Az egyiket már korábban említettük, de fontossága miatt nem árt újra hangsúlyozni, hogy a színezésnél figyelemmel kell lenni arra, hogy a számítógépen jól látható és egymástól jól megkülönböztethető színek a fekete-fehér skálára való áttéréskor (nyomtatás) azonos vagy nagyon hasonló árnyalatokat vehetnek fel. Ezért célszerű az ábrák elkészítésekor azonnal fekete-fehér változatot készíteni, és azon is jól megkülönböztethető árnyalatokat választani. A másik hiba akkor követhető el, ha a térkép nem közismert területet ábrázol. Ha – és ez a leggyakoribb – kiinduló pontunk Magyarország térképe, akkor az minden további megjegyzés, tájékoztatás nélkül érthető bárki számára. Ha azonban

például egy megye (vagy nem magyaroknak szóló publikációban az ország) térképe a kiinduló pont, akkor az alakzat ismeretlen lehet, ezért a felhasználót célszerű tájékoztatni annak elhelyezkedéséről, méreteiről, meg kell jelölni egy-két tájékozási pontot, esetleg mellékelni lehet egy nagyobb léptékű térképet.

Minőségi ismérvek szerinti megoszlások (például pártpreferenciák, foglalkozások, ágazati megoszlások stb.) ábrázolására legcélszerűbb a kördiagram használata.

A 9. ábra Magyarország energiamérlegének egyik elemét, az energiahordozók szerinti összetételt mutatja. Az ábráról élesen kiténik a szénhidrogének vezető szerepe, valamint az, hogy az atomenergia nem elhanyagolható részt képvisel a felhasználásban. (Ez utóbbi súlya jóval nagyobb lenne, ha a megoszlást csak a hazai termelésű energiahordozókra számítanánk.)

9. ábra. Az egyes energiahordozók százalékos aránya az összes felhasználásban, 1999



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

A minőségi ismérvek esetén teljesen felesleges a 3D ábrázolás, a harmadik dimenzió ugyanis nem hordoz információt. A 9. ábra helyett tehát nem ajánlott annak 3D formájú változata.

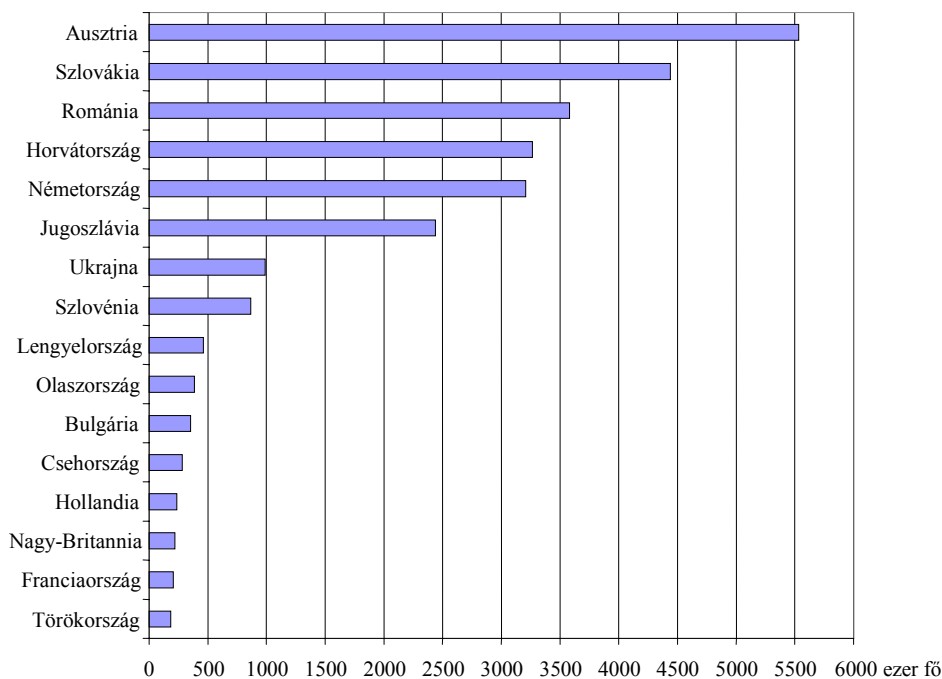
A minőségi ismérvek szerinti ábrázolásnál nagyon gyakori a helytelen típusválasztás. Mivel a minőségi ismérvek változatai általában nem rendezhetők egyértelmű sorrendbe, a kördiagram éppen ezt a semleges kiindulópontot tükrözi. Az erre a célra gyakran alkalmazott oszlopdiagram, ha azt egy szokásos derékszögű koordináta-rendszerbe helyezzük, balról jobbra értelmezhető, így eleve sugall valamiféle sorrendet. Ez pedig sérti a semleges ábrázolás elvét. Minőségi ismérvek esetén tehát alapesetben a kördiagramot részesítjük előnyben az oszlopdiagrammal szemben. Megjegyzendő, hogy a pontosabb értékelés érdekében olykor az egyes körszeletekhez hozzárendeljük a megfelelő megoszlási viszonyszámot (többnyire százalékos formában). Ez látható a 9. ábra kördiagramján is.

Természetesen lehetnek kivételek is. Ha az ismérvváltozatok száma nagy, a kördiagram áttekinthetetlenül sok szeletből áll; ilyen esetekben elkerülhetetlen az oszlopos ábrázolás. Ekkor azonban – éppen az említett feleslegesen sugallt sorrendiség elkerülése érdekében – inkább a fekvő oszlopos ábrázolás (szalagdiagram) javallott, ahogy az a 10. ábrán is szerepel, bár a megoszlásokat ez sem mutatja jól.

A 10. ábrával kapcsolatban megjegyezzük, hogy a nagyság szerinti rendezés nem kötelező, és a minőségi ismérvek szerinti megoszlás általában nem is használja, olykor mégis célszerű és informatív, hiszen, ha sok és

nehezen áttekinthető ismérvváltozattal rendelkezünk, akkor a sorba rendezés segít a jelenség jobb felismerésében. A bemutatott példában azonnal látható, hogy mely szomszédos országok adják az idelátogatók legnagyobb hányadát. (Ugyanez a jelenség kördiagramon ábrázolva áttekinthetetlenül töredezett lenne.)

10. ábra. A Magyarországra néhány európai országból érkezett látogatók száma 1999-ben



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

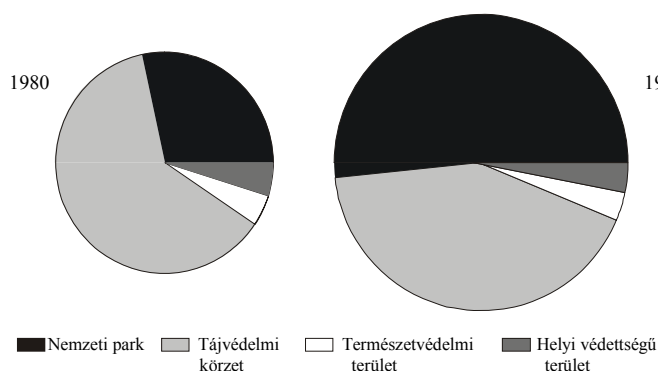
Amennyiben minőségi ismérvek szerinti megoszlások térbeli vagy időbeli összehasonlítására kerül sor, nagyon fontos, hogy a területekkel történő ábrázolás esetén az ábrák, illetve ábrarészek területeinek arányosaknak kell lenniük az ábrázolt jelenségekkel. Ha például kördiagrammal ábrázolunk két megoszlást (például két terület egység összehasonlításában), akkor a körök területe (tehát a sugarak négyzetgyöke) arányos kell legyen az ábrázolni kívánt gyakoriságokkal, illetve megoszlási viszonyszámokkal.

A 11. ábra két kördiagramja Magyarország védett területeinek megoszlását mutatja be, összehasonlítva az 1980-as és az 1999-es helyzetet. Az 1980-as megoszlás, azt mutatja, hogy a legnagyobb részt a tájvédelmi körzetek tették ki, míg a nemzeti parkok, bár területük nem elhanyagolható méretű és arányú volt, messze elmaradtak az előzőtől. A változás a vizsgált időszakban jól látható az ábrákon: az összes védett terület közel kétszeresére nőtt (a második kör sugara körülbelül 1,4-szerese az első körének), és az arányok is igen jelentősen megváltoztak: 1999-ben a védett területek több mint fele nemzeti park, míg a kisebb jelentőségű területek (természetvédelmi területek, helyi védettségű területek) aránya lényegesen csökkent. A 11. ábra a területarányos ábrázolás fontos példáját adja.

Ilyen esetekben az *osztott oszlopdiagram* a kördiagram vetélytársa lehet, amelynél az egyes oszlopok területe arányos az összehasonlító jelenségekkel, az egyes oszlopok

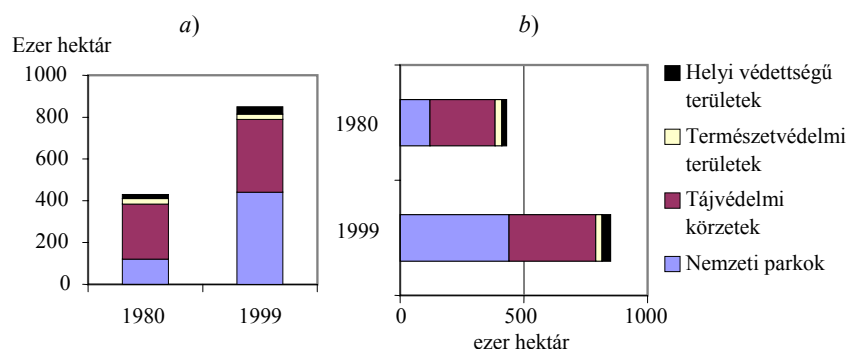
felosztása pedig az ismérvváltozatok arányait tükrözi. Ekkor az oszlopok függőlegesen, de akár vízszintesen is elrendezhetők. A 11. ábra ezen változatai a 12. ábrán láthatók.

11. ábrák. Védett területek megoszlása Magyarországon 1980-ban és 1999-ben



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

12. ábra. Védett területek Magyarországon 1980-ban és 1999-ben



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

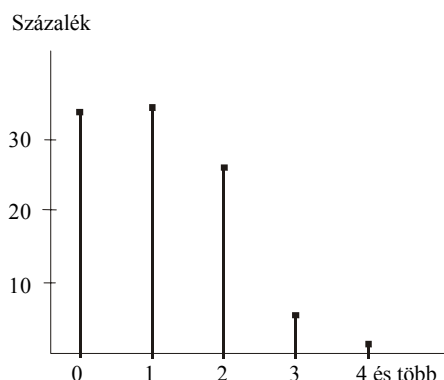
A 12. ábra két változata közül bármelyik választható, ám – tekintve, hogy itt az összehasonlítás idődimenzióban (is) történik – az a) ábra jobban mutatja az időbeli egymásutániségot. Megjegyezzük, hogy oszlopdiagramok esetén olykor alkalmasabb az az ábrázolás, amelyik az oszlop magasságát rögzíti (100 százalékban), és a növekedést az oszlopok eltérő szélességével fejezi ki. Természetesen ekkor is figyelemmel kell lenni a területarányosság követelményére. Ezt az ábrázolási formát a gyakran használt szoftverek sajnos nem támogatják.

Mennyiségi sorok esetén leggyakoribb a nagyság szerinti megoszlás ábrázolása. Amennyiben a csoportosító ismérv diszkrét, és kevés változata van,² célszerű forma a *pálcikadiagram*, amelyet a 13. ábra mutat be.

² A kevés ismérvváltozat annyit jelent, hogy az ismérvváltozatok könnyen áttekinthetők, ha úgy tetszik, egy táblában vagy egy ábrán felsorolhatók. Gyakorlatilag ez 15-20-nál nem több változatot jelent. Ilyen például a családok megoszlása gyermekszám szerint, a lakásállomány megoszlása szobaszám szerint stb. Diszkrét, de sok változattal rendelkező ismérv például a kor, ha azt egész években mérjük, hiszen a lehetséges változatok száma mintegy 100, ami már egy táblázatban vagy ábrán nem áttekinthető. A kevés vagy sok természetesen ez esetben sem nélkülöz bizonyos szubjektív megítélést.

A 13. ábra a magyar családok gyermekek száma szerinti megoszlását mutatja. Mivel a gyermekszám csak egész érték lehet, és az ismérvváltozatok száma kicsi (a 4 és több gyermekes családok aránya az összes családon belül nem éri el a 2 százalékot, ezért az ábrázolás szempontjából ezt egy ismérvtétknek tekinthetjük). A diszkrét értékek indokolják a pálcikaábrázolást, ugyanakkor tekintve az ábra egyszerű szerkezetét és a kevés vonalat, nem teszi szükségessé a pontok összekötését, anélkül is jól látszik az eloszlás jellegzetessége: erőteljesen balra ferdülő alakja.

13. ábra. A családok gyermekszám szerinti megoszlása, 1996



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Megjegyezzük, hogy a gyakorlatban ilyen esetekben is lényegesen elterjedtebb az oszlopdiagram használata. Mivel ilyenkor a területtel való ábrázolás indokolatlan, feleslegesen bonyolultnak, ezért kerülendőnek tartjuk a hisztogramos ábrázolást. Nem szokták a pálcikák végpontjait összekötni, mert ez ellentétes a jelenség diszkrét természetével (például 2,35 szobás lakás nem értelmezhető). Ugyanakkor, ha az eloszlás jellegzetességeit szeretnénk folytonos szakaszokkal kiemelni (ami a poligonnal való ábrázolásnak felel meg), a pontok összekötése nem kifogásolható.

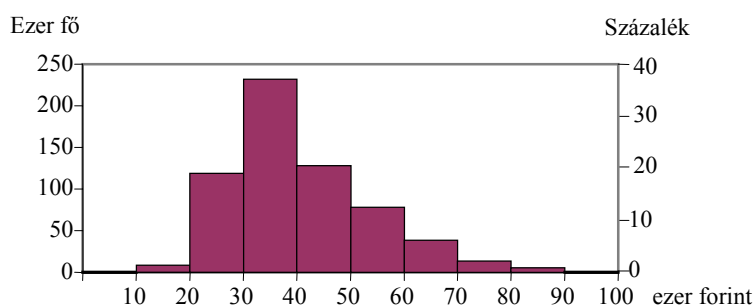
A társadalom és gazdaság leírásakor a *menyiségi ismérvek szerinti megoszlás* ábrázolása a leggyakoribb olyan esetekben, amikor a csoportosító ismérv folytonos, illetve diszkrét ugyan, de sok ismérvváltozattal rendelkezik. Ilyen például a kor, a jövedelem, a testsúly stb. szerinti megoszlás. Ezt a jelenséget oszlopdiagrammal, más néven *hisztogrammal* ábrázoljuk. A hisztogramos ábrázolás esetén célszerű arra törekedni, hogy az oszlopok hézagmentesen illeszkedjenek egymáshoz. Amennyiben ún. közölt osztályhatárokat képezünk, ennek nincs semmi akadálya. A hézaggal illeszkedő oszlopokat kerüljük még akkor is, ha egyes szoftverek alapértelmezésben ezt támogatják. Ez ugyanis a folytonosság hiányára utal, ugyanakkor megtöri a hisztogram által az eloszlásról közvetített képet.

A hisztogram az idősoros vonaldiagram mellett talán a leíró statisztika leggyakrabban használt ábrája, ezért alkalmazásával kapcsolatban több megjegyzést kell tennünk. Az első arra vonatkozik, hogy hisztogrammal mind abszolút nagysággal rendelkező megoszlási viszonyszámokat, mind ezek százalékos változatát, a relatív gyakoriságot kifejező megoszlási viszonyszámokat egymással egyenértékűen lehet ábrázolni, sőt ez a két ábrázolás azonos hisztogramokat eredményez. Ezért ebben az esetben nem tekinthető hibának

az, ha a két függőleges tengelyen más és más skála szerepel. Ez a kétcsöví megoldás ekkor ugyanis azonos ábrát eredményez, hiszen a relatív gyakoriságok arányosak az abszolút gyakoriságokkal. Ha a bal oldali tengelyen az abszolút, a jobb oldalin a relatív gyakoriságokat jelöljük, akkor a látszólag két ábra egybeesik, ezért ebben a speciális esetben kifejezetten jó a két különböző skála feltüntetése, hiszen ez ugyanannak a jelenségnek két egyenértékű vetületét mutatja meg.

A 14. ábra hisztogramja a nyugdíjas férfiak számának öregségi nyugdíjak nagysága szerinti megoszlását mutatja. Az egymáshoz hézagmentesen illeszkedő oszlopok arányai függetlenek a két függőleges tengely eltérő skálájától. Az ábra jellegzetes jövedelemeloszlást mutat: a változó (nyugdíj), csak pozitív értékeket vehet fel a módusz a mediántól balra helyezkedik el, azaz a közepesenél kisebb nyugdíjak környezetében sűrűsödnek az értékek; a jellemző nyugdíjak a 30 ezer és 40 ezer forint közötti intervallumban találhatók, az eloszlás balra ferde, jobbra elnyúló. (A jövedelmek eloszlása is hasonló jellegzetességeket mutat, de erősebb ferdeséggel, a nyugdíjak ugyanis kevésbé egyenlőtlenül oszlanak meg, mint a jövedelmek.)

14. ábra. A férfi népesség megoszlása az öregségi nyugdíjak nagysága szerint 2000. január 1.



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

A hisztogrammal való ábrázolás fontos kérdése a területarányosság biztosítása, amit a minőségi ismérvek szerinti megoszlás kapcsán már vizsgáltunk. Gyakori ugyanis, hogy az ábrázolás alapjául szolgáló osztályközös gyakorisági sor nem egyenlő osztályközökre oszlik. Ilyen eset gyakran előfordul akkor, amikor szélsőségesen nagy vagy kicsi értékek nem elhanyagolható számban fordulnak elő a sokaságban, de akkor is, amikor a szélső osztályközök nyitottak. Ezekben az esetekben az ábrázolás korrektsége megköveteli az arányosítást, azaz azt, hogy a gyakoriságokat (relatív gyakoriságokat) egyenlő osztályközökre számítsuk át. Ha ezt nem tesszük meg, félrevezető ábrát kapunk, a megfelelő korrekció után viszont az ábra már a valóságos jelenséget tükrözi.

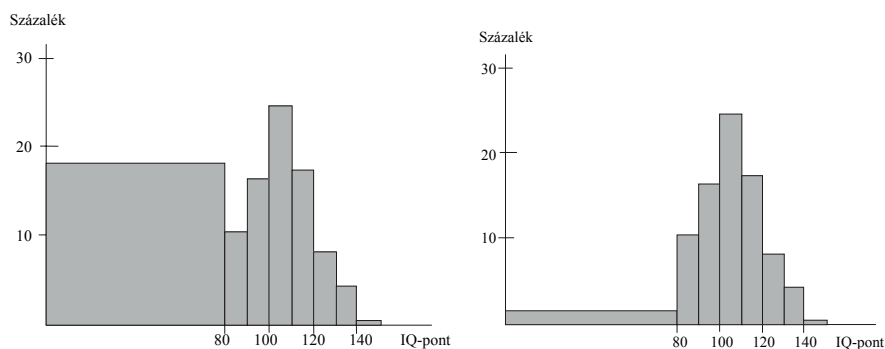
A 15. ábrák egy, a fiatalok körében végzett pszichológiai felvétel stilizált eredményeit mutatják. A vizsgálatban IQ-tesztek eredményei alapján csoportosítottuk a népességet az alábbiak szerint.

Egy népességcsoport IQ-pontok szerinti megoszlása

IQ-pont	Százalék	IQ-pont	Százalék	IQ-pont	Százalék
-80	18,2	101-110	24,7	131-140	4,1
81-90	10,8	111-120	17,3	141-150	0,5
91-100	16,2	121-130	8,2	Összesen	100,0

Ha a hisztogramos ábrázolást ezen megoszlási adatokkal végezzük, a módfelett félrevezető 15. ábra első grafikonját kapjuk: a 80 pont alatti értékek relatív gyakoriságát az egész intervallumra kivetítve az első intervallum nagy gyakorisága azt sugallná a felületes szemlélő számára (és ne feledjük, az ábrák a felületes szemlélőknek éppúgy szólnak, mint az elmélyült elemzőknek, sőt hatásuk az előbbi csoportra nagyobb), hogy a népességben jelentős hányadot képviselnek a debiliek.

15. ábra. Egy népességcsoport megoszlása IQ-pontok szerint



Forrás: saját adatgyűjtés és számítások.

Ez az ábrázolás félrevezető, hibás! A hiba oka az, hogy ez az ábrázolás nem veszi figyelembe a területarányosság követelményeit, és az első, a többi kategória osztályköz-hosszát nyolcszorosan (!) felülmúló osztály teljes hosszára azt a relatív gyakoriságot tekintti érvényesnek, ami csak a jóval kisebb osztályközökre vonatkozatható. Ezért, a területarányosság szem előtt tartásával úgy lehet helyesen ábrázolni a jelenséget, hogy az első osztályköz gyakoriságát korrigáljuk, ami esetünkben annyit jelent, hogy a hisztogram magasságát az eredeti nyolcadrésénél húzzuk meg.

Az így kapott 15. ábra második grafikonja már korrigálja az említett aránytalanságot, bár feltételezése (nevezetesen az, hogy a 80 pont alatti értékek egyenletesen oszlanak meg 0 és 80 között) félrevezető. Mégis, ha más információk nincsenek, ezt az ábrázolást kell választani, mert ez felel meg legjobban az ábrázolás általános elveinek.

Ehhez a példához két megjegyzés kívánkozik. Egyrészt a valóságban ritkán jelentkeznek ennyire élesen a problémák; ez a példa konstruált, ezért mutat ilyen éles különbségeket. Másrészt sajnálatos módon a gyakran használt szoftverek (Excel, SPSS) nem támogatják ezt a fajta területarányosítást, ezért a mechanikusan készített gépi ábrák gyakran hibásak.

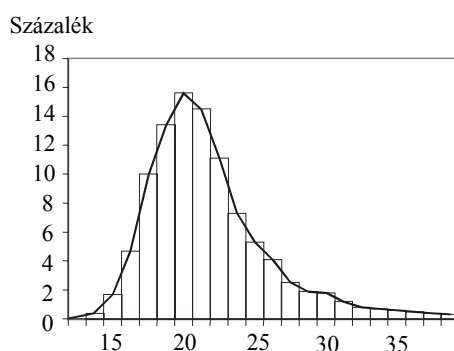
A hisztogrammal való ábrázolás annyira fontos a statisztikában, hogy más esetekben (például a korábban említett tartamidősorok bemutatásánál) kerülni kell a hasonló ábrázolást annak érdekében, hogy fenntartsuk a hisztogram kiemelkedő és központi jelentőségű szerepét. Ezt a szerepet kiemelik a következő további tulajdonságok.

A hisztogramok oszlopainak középpontjait összeköthetjük egyenes szakaszokkal. Ez megengedett művelet, és nem az osztályokon belüli folytonosságra utal, hanem önálló ábraként a nagyság szerinti eloszlás jellegét domborítja ki. Ezt az ábrát a statisztika – kiemelkedő jelentőségére való tekintettel – külön névvel illeti: ez a *poligon*. Attól függően,

hogy a poligon alapjául szolgáló hisztogram gyakoriságokra vagy relatív gyakoriságokra épül, gyakorisági vagy relatív gyakorisági poligonról beszélünk. Egyazon jelenség esetén a kettő arányaiban tökéletesen megegyezik. A relatív gyakorisági poligon fontos tulajdonsága, hogy a görbe alatti terület egységnyi, amit a hasonló háromszögekre vonatkozó tételek alapján igen könnyű belátni. A relatív gyakorisági poligon valójában nem más, mint a valószínűség-számításban fundamentálisnak minősülő *empirikus sűrűségfüggvény*.

A 16. ábra egy gyakorisági poligont mutat be. Az ábra egy jellegzetesen balra ferde eloszlás hisztogramjából indul ki, és erre rajzolja a poligont, amely – tekintve, hogy a hisztogram relatív gyakoriságokra épül – relatív gyakorisági poligon lesz. A vizsgált jelenség ez esetben a magyar népesség elhízása, az alkalmazott mutató pedig a BMI (Body Mass Index), amely a testsúly és a testmagasság arányából képzett egyszerű mutató: általános megítélés szerint 16 alatti értéke soványságra, 25 feletti túltápláltságra, 30 feletti kóros túlsúlyra utal. A poligon az eloszlás szabályos alakja folytán jól kirajzolja a lognormális eloszlást, és a sűrűségfüggvény megalapozója lesz. Az ábráról jól látható a jellegzetes bal oldali aszimmetria, azaz az, hogy a magyar népességben magas a túlsúlyosok aránya.

16. ábra. A sorköteles fiatalok BMI-index szerinti megoszlása

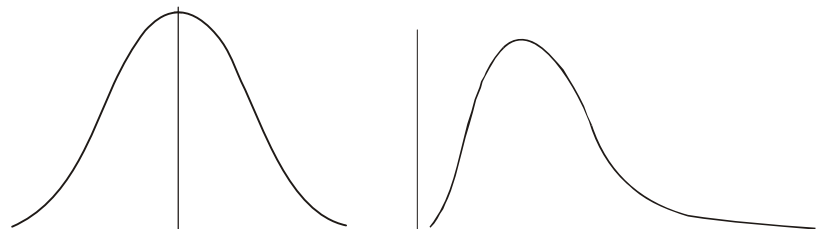


Forrás: Joubert – Gyenis; 2001.

Ha a gyakorisági poligon esetén az osztályközök tetszés szerint kicsivé tehetők (a felosztás korlátlanul sűrítethető), és a rendelkezésre álló megfigyelések száma megengedi, hogy az így megnövekedett számú osztályba is elegendő számú elem essék, akkor az esetek jó részében a poligon egyre közelebb jut egy folytonos görbéhez, amelynek elnevezése *gyakorisági görbe*. Relatív gyakoriságok esetén a gyakorisági görbe nem más, mint a valószínűség-számításból ismert *sűrűségfüggvény*. Mivel több társadalmi–gazdasági jelenség jól közelíthető valamely ismert valószínűség-eloszlás segítségével, ezek sűrűségfüggvényeit, illetve gyakorisági görbéit gyakran idézik a hisztogramok mellett vagy helyett.

A 17. ábra egy normális eloszlás, a 18. ábra egy lognormális eloszlás sűrűségfüggvényét vázolja. A normális eloszlás a leggyakrabban előforduló eloszlástípus, amely a méreteltérések, a hibák, valamint az egyenként számba nem vehető tényezők eredőjeként adódó változók jellemző törvényszerűségeit írja le. A lognormális eloszlás jellegzetesen balra ferde eloszlás; a gazdasági gyakorlatban leginkább a jövedelmek eloszlásának leírására használják, de más, nagyság szerinti eloszlások jellemzésére is alkalmas. (A 15. ábra adatai kellő pontosságú mérés és osztályozás esetén normális, a nyugdíjnak a 14. ábrán bemutatott eloszlása vagy a BMI-index 16. ábrán vázolt eloszlása lognormális modellt követ.)

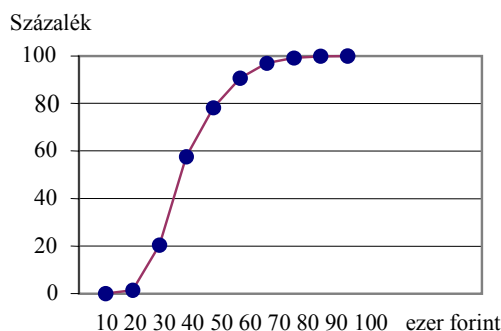
17. ábra. A normális eloszlás sűrűségfüggvénye 18. ábra. A lognormális eloszlás sűrűségfüggvénye



Még mindig a hisztogramból kiindulva képezhetők kumulált eloszlások, melyeket ugyancsak egymáshoz hézagmentesen illeszkedő oszlopdiagramokkal illusztrálunk. Ha ezen oszlopok tetejének felezőpontjait egyenes szakaszokkal összekötjük, akkor egy újabb ábrát, az ún. *ogivát* kapjuk. Amennyiben gyakoriságokra épített ogiva esetén végezzük el a felosztás említett sűrítését, akkor kumulált gyakorisági görbét kapunk, ha pedig minderre relatív gyakoriságokból kiindulva kerül sor, akkor az eredmény a kumulált relatív gyakorisági görbe.

A 19. ábrán az öregségi nyugdíjasok megoszlását bemutató 14. ábra hisztogramjából kiindulva állítottuk elő az ogivát, amely szabályos alakja folytán igen jól közelíti a kumulált relatív gyakorisági görbét.

19. ábra. A férfi öregségi nyugdíjasok megoszlásának ogivája



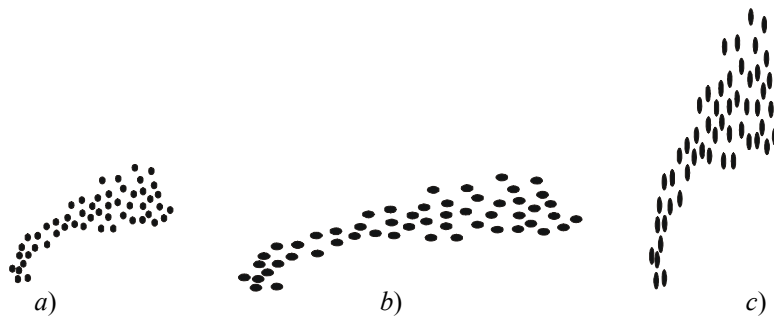
Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

A kumulált relatív gyakorisági görbe fogalmilag megegyezik a valószínűség-számítás eloszlásfüggvényével. Ez az ábratípus kumulált jellegénél fogva nem mutat jól értelmezhető jellegzetességeket, ezért ritkábban használjuk. Mindenképpen meg kellett azonban említeni, hiszen a valószínűség-számítás megalapozásában döntő szerepe van.

Az eddig bemutatott ábrák voltaképpen egydimenziós feladatokat mutattak be, jóllehet például két különböző jelenség időbeli vagy térbeli megoszlása már többdimenziós jelenség. A jellegzetes többdimenziós feladatok azonban általában két vagy több mennyiségi ismérv kapcsolatából indulnak ki. Legegyszerűbb esetben ezt a kapcsolatot pontdiagrammal írjuk le, ahol a pontfelhő elhelyezkedése, alakja, sűrűsége és egyéb jellemzői mutatják a két változó kapcsolatát. Ez esetben is nagyon kell ügyelni arra, hogy a skálák megváltozása lényegesen befolyásolhatja a pontfelhő jellemzőit, ezért ajánlatos többféle

skálát kipróbálni, és azt közölni, amelyik a leginkább semlegesnek tűnik. (Lásd a 20. ábrát.) Ez természetesen jelenthet némi szubjektív megítélést.

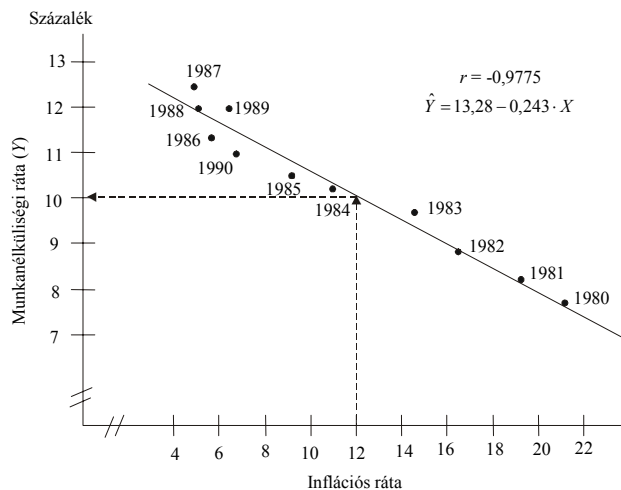
20. ábra. A kétdimenziós pontfelhők elhelyezkedése



A 20. ábra ugyanannak a pontfelhőnek az egyszerű, függőleges vagy vízszintes nyújtásából származnak. Minden különösebb megfontolás nélkül látható, hogy ezek az alakzatok mind más minőségi megállapítást tesznek lehetővé a két változó kapcsolatára vonatkozóan. Ezek közül az a) ábra látszik a legkevésbé szélsőségesnek.

Amennyiben nem elégszünk meg azzal, hogy pontfelhővel ábrázoljuk a kapcsolatot, hanem a feltárt például regressziós összefüggéseket is be akarjuk mutatni, akkor ezt folytonos vonallal célszerű megtenni. A folytonosság egyrészt kiemeli, hangsúlyozza a tendenciát, másrészt arra is utal, hogy a nem megfigyelt helyeken is lehet a függvényt értelmezni (folytonos változók feltételezésével). Diszkrét esetben ez természetesen nem tehető meg, de a tendencia jelzésére a folytonos függvényt ekkor is be lehet rajzolni az ábrába. A megfigyelési pontokat olykor a megfigyelés helyére, idejére stb. vonatkozó utalásokkal is el lehet látni.

21. ábra. Az infláció és a munkanélküliség kapcsolata Olaszországban



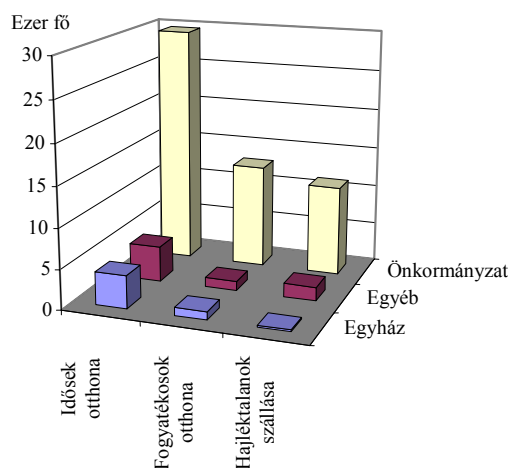
Adatforrás: Külkereskedelmi statisztikai évkönyv, 1992 (1993). Központi Statisztikai Hivatal, Budapest.

A 21. ábra egy korrelációs és regressziós kapcsolatot mutat be. A vizsgált összefüggés egy makrogazdasági feladat: az inflációs ráta és a munkanélküliségi ráta kapcsolata az olasz gazdaság idősoros adataira építve. A pontfelhő egyes elemeit az évszámra utaló azonosítókkal láttuk el, és szerepel az ábrában a korrelációs együttható, valamint a becült regressziós egyenes egyenlete is. A diszkrét megfigyelésekre illesztett folytonos egyenes utal arra, hogy a megfigyelési intervallumon belüli tetszőleges inflációs rátához – akár grafikusan is – meghatározható a megfelelő becült munkanélküliségi ráta. Ilyen grafikus becslést a 12 százalékosnak feltételezett inflációból kiindulva a szaggatott vonal mutat be. Az ábrával kapcsolatban felhívjuk a figyelmet arra, hogy mindkét tengelyt megszakítottuk, amit jelöltünk. Ezt az azért fontos megemlíteni, mert ez is befolyásolja az ábra optikáját; ha nem szakítjuk meg a tengelyeket, a koordinátatengelyekhez képest más helyzetű pontokat és egyenest kapunk.

Kétdimenziós megoszlások esetén természetesen jól használhatók a számítógépes csomagok által felajánlott háromdimenziós ábrák. Ekkor a vízszintes síkot határoló tengelyeken a két ismérv változatai szerepelnek, és a kombinált gyakoriságok a harmadik dimenzióban jelennek meg. Ez az ábrázolás valójában egy kétdimenziós kombinációs tábla gyakoriságainak vagy relatív gyakoriságainak axonometrikus ábrázolása. Ez esetben az ismérvek egyaránt lehetnek mennyiségiek és minőségiek. Ezt az ábrázolást *3D hisztogramnak* nevezzük.

A 22. ábra a szociális intézményekben gondozottak számának kombinált megoszlását mutatja be egyrészt az intézmény jellege, másrészt a fenntartó szervezet szerinti bontásban. Látható, hogy mindkét csoportosító ismérv minőségi, így valójában egy asszociációs táblát ábrázolunk. A jellemző tendenciák, nevezetesen az önkormányzati tulajdon túlsúlya, valamint a jelleg szerinti megoszlások arányainak hasonlósága jól látható az ábrán.

22. ábra. A szociális intézményekben gondozottak száma
intézménytípus és fenntartó szerint, 1999

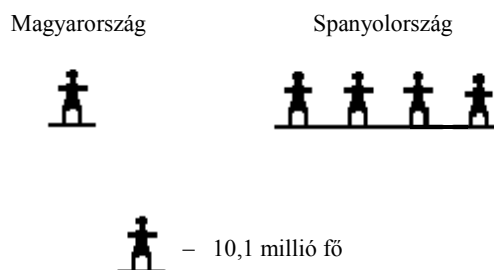


Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Végezetül még egy ábratípus, amely talán a legegyszerűbb és legszemléletesebb, ugyanakkor természetesen a legkevésbé rugalmas, és a legkevésbé alkalmas a jelenségek árnyalt ábrázolására. Ez a *piktogram* vagy más néven *piktográf*. A piktogram a legszélesebb közvélemény számára közvetíthet egyszerű statisztikai információkat valamilyen fi-

gurális ábrázolás révén. Mivel síkidomokról van szó, elvben itt is érvényes a területarányos ábrázolás követelménye, azaz az idomok területi arányainak meg kell felelniük az ábrázolandó jelenség arányainak. Ennek biztosítása nem egyszerű feladat, hiszen gondoljunk csak arra, hogy például egy embert ábrázoló képet hogyan lehet úgy felnagyítani, hogy területe mondjuk négyszerese legyen egy másik hasonló ábrának. (Ha például Magyarország és Spanyolország népességének arányát kívánjuk szemléltetni, akkor ilyen feladattal állunk szemben.) Ezért ilyen esetekben inkább az azonos méretű idomok többszörözését használjuk, bár ekkor a nem egész számú arányok okozhatnak problémát. Ez a feladat tehát például a 23. ábrán látható módon oldható meg.

23. ábra. Magyarország és Spanyolország népessége, 1999



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

A pictogram – jóllehet szerepét a tömegtájékoztatásban nem lehet elhanyagolni – csak egyszerű jelenségek elnagyolt ábrázolására alkalmas. Ekkor is figyelemmel kell lenni azonban a helyes ábrázolást elősegítő elvekre és szabályokra.

ÖSSZETETT ÁBRÁK

Az eddigiekben a gyakorlatban legtöbbször előforduló legegyszerűbb ábratípusokat, azok helyes és helytelen alkalmazását mutattuk be. A statisztikai jelenségek grafikus ábrázolásának lehetőségei azonban korántsem merülnek ki ezekkel. A továbbiakban azokkal az ábrázolási módokkal foglalkozunk, amelyek az elméletben jól ismertek, számítógépes csomagok többnyire támogatják őket, ugyanakkor alkalmazásuk szórványos. Tekintve, hogy az ábrák, ábratípusok száma szinte korlátlan, nem törekszünk teljességre, csupán a valóban egyszerű és jól használható ábratípusokra mutatunk rá.

A mennyiségi sorok elemzésének eszközei közül most két olyan ábratípust mutatunk be, melyek egyszerűen, mégis jó hatásfokkal alkalmazhatók gyakorisági sorok leírásakor, s melyek alkalmazását feltétlenül ajánlani tudjuk, hiszen ezek a népszerű statisztikai programcsomagoknak is részét képezik.

Az egyik, amelyet abban az esetben célszerű alkalmazni, ha a sokaság viszonylag kevés elemből áll, és az elemek külön-külön (tehát nem csak csoportosított formában) ismertek. A *leaves ág* (stem and leaves) elnevezésű ábra a sokaság elemeinek a vizsgált ismérv szerinti felsorolását tartalmazza megfelelő elrendezésben. Ez az elrendezés annyit jelent, hogy az ismérvértékek első számjegyét (helyértékét) egy vonal elé kiemeljük, és

az ismérvtételeket már csak az első számjegy sorába írott második, harmadik stb. számjegyeikkel jellemezzük.

A 24. ábrán a 2001. tanévben a Budapesti Közgazdasági és Államigazgatási Egyetem három csoportjába tartozó 99 hallgató által 2001. októberben írt statisztikadolgozat eredményei szerepelnek. (A dolgozattal legkevesebb 0 és legfeljebb 50 pontot lehetett szerezni.)

24. ábra. Leveles ág ábra 99 hallgató statisztika dolgozatának pontszámairól

Szár	Levelek
0	6, 7, 9,
1	0, 1, 1, 2, 2, 2, 2, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9,
2	0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 5, 6, 6, 6, 7, 7, 7, 8, 8, 8, 9, 9, 9,
3	0, 0, 0, 0, 1, 1, 2, 3, 3, 4, 4, 4, 6, 6, 8, 8, 8, 9, 9, 9,
4	0,

Forrás: saját számítások.

Az ábra elemei szerint a sokaságban volt egy 6, egy 7 és egy 9 pontos dolgozat, az 1-gyel kezdődő, azaz a 10-19-es pontosak között volt egy 10, két 11, négy 12 pontos dolgozat stb. A sor végén egy 40 pontos dolgozatot mutat az ábra.

Ez az ábra azért jó, és azért válik ilyen feladatok esetén egyre népszerűbbé, mert

- megadja a teljes sokaságot (valamennyi elemét), tehát munkatáblaként lehetőséget ad az egyedi információk hasznosítására;
- megadja a rangsort, azaz a sokaság elemeit sorba rendezi, lehetővé téve a sokaságra vonatkozó rendezett mutatók (például kvantilisok) könnyű és gyors meghatározását;
- alakja kiemeli az eloszlás jellegzetességét, az ábrát 90 fokkal balra forgatva ugyanis egy hisztogram bontakozik ki, amely – mint mondtuk – a nagyság szerinti eloszlás legjellemzőbb ábrája (az ábrát így már közlési táblaként is lehet alkalmazni).

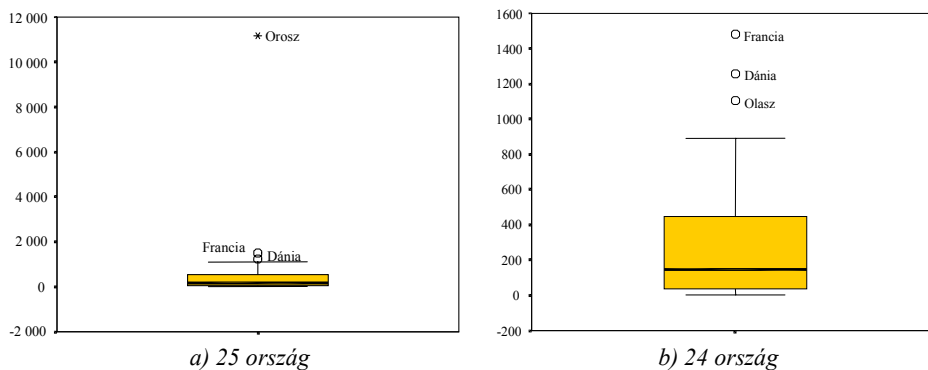
A mennyiségi sorok árnyaltabb leírására szolgáló ábrák közül a *box-plot* (elfogadott magyar elnevezése nincsen) használatát javasoljuk. A box-plot a mennyiségi eloszlások néhány egyszerű jellemzőjét, így

- várható értékét (átlagát),
- kvartiliseit (beleértve természetesen a mediánt is),
- nyesett terjedelmét

mutatja egyetlen ábrában, és emellett külön meg is jelöli a kiugró (outlier) értékeket.

A 25. ábrán 25 ország 1996. és 1997. évi munkabeszüntetései számát vizsgáltuk. Az *a*) ábra esetében az országok közt figyelembe vettük Oroszországot, ahol kiugróan magas volt 1997-ben a munkabeszüntetések száma. Ezért annak érdekében, hogy Oroszország is ráférjen az ábrára, igen kis léptéket kellett alkalmazni, így a többi ország adatai olyan szűk intervallumra sűrűsödtek, hogy a box-plot alig értékelhető. (Az ábra árnyékolt területe szinte egy egyenessé zsugorodott össze.) Ha ellenben Oroszországot elhagyjuk (lásd a *b*) ábra), a többi vizsgált ország adata egy jól értékelhető dobozt rajzol ki, amelyből leolvasható a sztrájkok átlagos száma, mediánja, kvartilisei és nyesett terjedelme. Ekkor a kiugró értékek (mértékeikben az előzőnél jóval szerényebbek) Olaszország, Dánia és Franciaország adatai. A két ábra összehasonlítása önmagában is jól jellemzi Oroszország különállását ezen ismérv szerint.

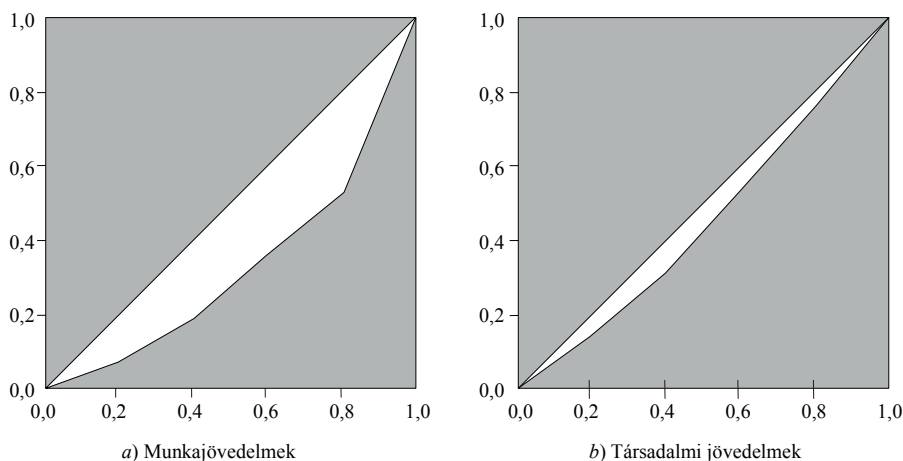
25. ábra. A munkabeszüntetések számának box-plot ábrája, 1996–1997



Adatforrás: Magyar statisztikai évkönyv, 1999. (2000) Központi Statisztikai Hivatal, Budapest.

Még mindig a mennyiségi soroknál maradva, elsősorban a gazdaságstatisztikában fontos feladat a *koncentráció* mérése. Koncentrációval (relatív koncentráció) valójában eloszlások egyenlőtlenségét vizsgáljuk. Közelebbről azt, hogy a sokasági értékösszeget hordozó elemek kumulált relatív megoszlása milyen viszonyban van a általuk képviselt értékösszeg kumulált relatív megoszlásával. A koncentrációt többé-kevésbé ismert mérőszámok (Gini-index, Herfindahl-index, entrópia és redundancia) mellett leggyakrabban a Lorenz-görbe elnevezésű, speciális szerkezetű ábrával (célábrával) lehet kimutatni. A Lorenz-görbe egy egységnyi oldalú négyzet oldalaira méri fel egyrészt a sokasági elemek kumulált relatív gyakoriságait, másrészt ezekkel párban az értékösszeg kumulált relatív gyakoriságait. Az így kapott pontokat összekötve egy szakaszokból álló vonalat kapunk, mely vonal és a főátló által bezárt terület nagysága mutatja a koncentráció mértékét.

26. ábra. Az 1999. évi jövedelemeloszlás koncentrációjának Lorenz-görbéi



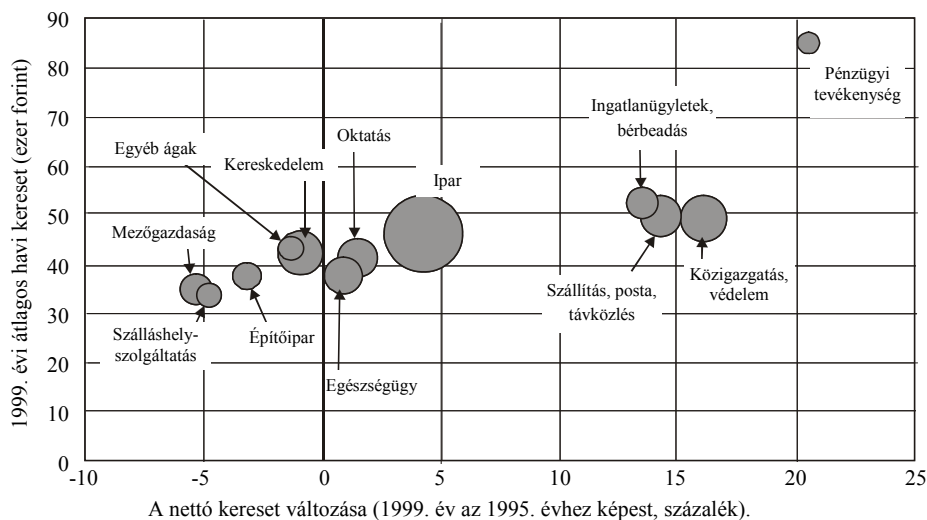
Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

A 26. ábrán az 1999. évi jövedelemeloszlás egyes elemeinek koncentrációját mutatjuk be. A vizsgálatot népességötödekre (kvintilisekre) végeztük, és az ábrák azt mutatják, hogy a népesség jövedelem szerint sorba rendezett ötödei az összes jövedelem hány százalékát szerezték meg. Az *a*) ábrán a munkajövedelmeket, a *b*) ábrán a társadalmi jövedelmeket mutatjuk be. Mivel a Lorenz-görbe a nem ármékolt terület nagyságával fejezi ki a koncentrációt, a két ábrából minden további számítás, elemzés nélkül jól látható, hogy a munkajövedelmek esetén a koncentráció (fehér terület) sokkal nagyobb, mint a társadalmi jövedelmek esetén. Ez így is van rendjén, hiszen a társadalmi jövedelmek egyik fontos szerepe az egyenlőtlenségek csökkentése. Meg kell ugyanakkor jegyezni, hogy ez az egyetlen mutató (ábra) nem elegendő a jelenség alapos elemzésére, hiszen az aggregált mutatószámot olykor egymásnak ellentmondó tendenciák alakítják.

Az egyszerűbb statisztikai jelenségek mellett természetesen a bonyolultabbak esetében is igen gyakran élünk az ábrázolással. Ilyenkor gyakori az, hogy egy-egy konkrét feladatra saját ábrát készítünk. A sokváltozós statisztika elvben nem a legjobb terep az ábrázolásra, hiszen az ábrázolás többnyire két dimenzióhoz van kötve, és csak ritka esetekben lehet 3 dimenziós feladatokat síkban érzékeltetni.

A 27. ábra erre a ritka esetre mutat példát. Az ábra az egyes gazdasági ágak helyzetét mutatja meg egyfelől az átlagos havi kereset, másfelől ennek növekedési üteme függvényében. Emellett azonban az objektumok (ágak) méretét is ábrázolja a szerző, mégpedig a nagysággal (foglalkoztatottak számával) arányos területű, megfelelő helyzetű kör segítségével. Így az ábra, amellett, hogy világos szerkezetű, egyszerre három ismérv szerint ad kombinált elemzési lehetőséget. Az ábra tartalmáról felesleges szólni, hiszen az önmagáért beszél.

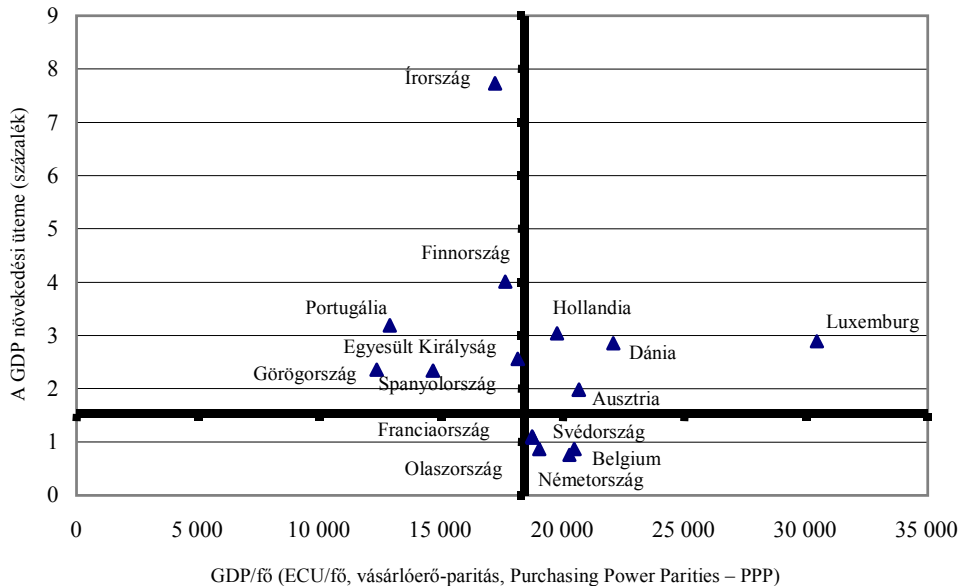
27. ábra. A nettó keresetek alakulása 1995 és 1999 között gazdasági áganként



Forrás: Magyarország 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Némiképp hasonló szerkezetű, de más, kifejezetten további elemzés céljából készült a 28. ábra, amely az Európai Unió országainak két ismérv szerinti jellemzésére szolgál. Ezen a szerző függőleges és vízszintes osztóvonalak segítségével elkülöníti a két ismérv (az egy főre jutó GDP és a GDP növekedési üteme) szerint átlag alatti és feletti országokat, és a későbbi elemzés arra szolgál, hogy az időbeli változások és az Unió feltételezett bővítésének hatásait ezen ábra (az ún. BCG-mátrix) segítségével szimulálja. Az időben bekövetkező változások, jelen esetben az esetleges bővülés, ugyanis megváltoztatják az osztóvonalak helyzetét, és ennek következtében az egyes uniós országok más kategóriába kerülhetnek át. Ennek kiinduló pontja a 28. ábra.

28. ábra. Az EU-országok elhelyezkedése az egy főre jutó GDP és a GDP növekedési üteme szerint



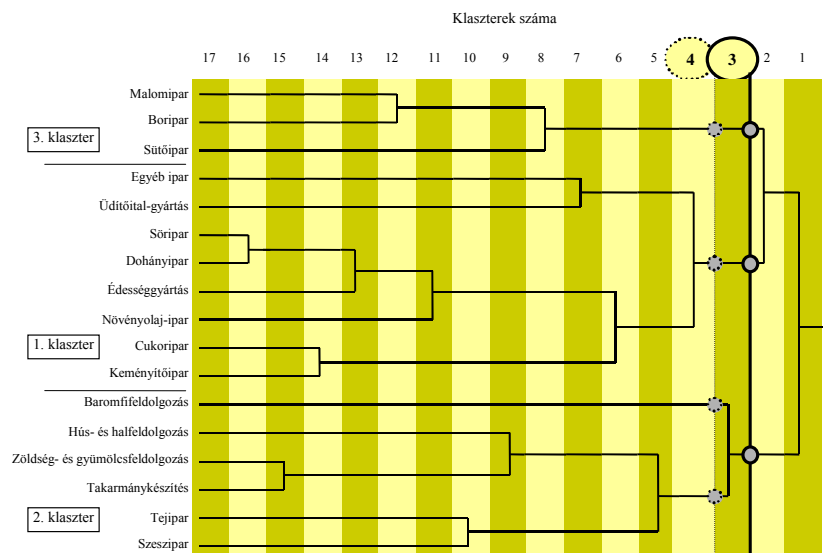
Forrás: Rappai, 2000.

A sokváltozós statisztika természetesen más ábrákat is használ: az objektumok elrendezését, távolságaikat, illetve közelségüket gráfok segítségével is megadhatják; a dimenzióredukáló eljárások nemritkán 2 vagy 3 dimenzióba tudják objektumaikat vagy megfigyeléseiket transzformálni, és ekkor már viszonylag egyszerű, a korábban bemutatottakhoz hasonló szerkezetű pontdiagramok segítségével a korábban ismertett ábrázolási technikákkal lehet dolgozni. Talán a legjellegzetesebb sokváltozós ábra mégis a *dendrogram*, amely a klaszteranalízis (automatikus osztályozás) egyik igen népszerű munkaábrája.

A 29. ábra egy dendrogramot mutat be. A feladat a magyar élelmiszer-gazdaság szakágazatainak osztályozása két ismerv: a külföldi tulajdon aránya és a koncentráció szerint. A dendrogram azt mutatja, hogy az egyes al-, illetve szakágazatok³ e két ismerv szerint hogyan kapcsolódnak fokozatosan (hierarchikusan) össze: az egymáshoz közel álló (vizsgált ismérvekben hasonló) szakágazatok már korán, a klaszterfa alsó szintjein összekapcsolódnak, és önálló csoportot hoznak létre, míg az egymástól lényegesen különböző objektumok (al-, illetve szakágazatok) csak a hierarchikus egyesítés magasabb szintjén, végső fázisaiban találkoznak. Így jól kirajzolódó csoportot alkot a söripar, az édesipar, a dohányipar, az üdítőital-gyártás, a növényolaj- és a cukoripar. Ezekre jellemző a nagy koncentráció, a nagyarányú külföldi részesedés és a gyorsan modernizálódó technológia. A másik jól elkülönülő csoportot egyebek között a tejipar, a szeszipar, a hús- és halfeldolgozás alkotják, amelyekre az alacsony koncentráció a jellemző, míg a harmadik csoportot a hagyományosnak minősülő, kevésbé koncentrált, tradicionális, kevésbé termelékeny technológiával dolgozó, jórészt hazai kézben levő alágazatok (malomipar, sütőipar, boripar) alkotják.

³ Ez az elemzés kifejezetten az élelmiszer-gazdaság egyes kérdéseit vizsgálta, ezért alkalmazott bontása ehhez igazodik: alapvetően alágazatokat vizsgál, de esetenként szakágazati mélységig is lemegy.

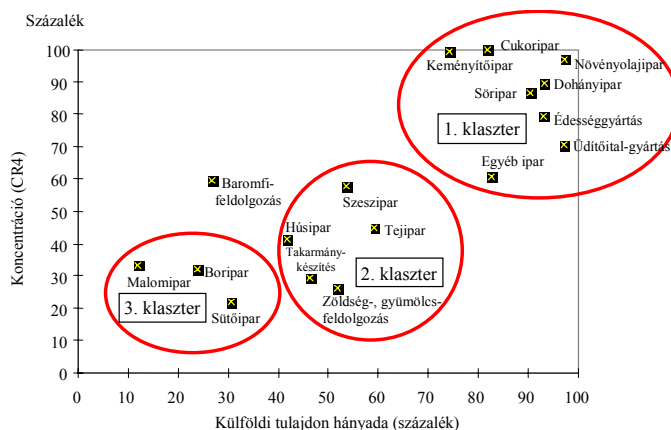
29. ábra. A magyar élelmiszeripari al-, illetve szakágazatok osztályozása



Forrás: Jansik; 2000.

A dendrogram lehetőséget ad az elemzőnek arra, hogy különböző számú csoportok esetén a jellemző összekapcsolódásokat, sűrűsödéseket könnyen számba vegye, az osztályozást több ismérv együttes figyelembevételével objektív mérce alapján elvégezze. A dendrogramot kiegészítheti egy olyan pontdiagram, amelyik a két vizsgált (vagy több ismérv esetén a két legfontosabb valódi vagy mesterséges) változó terében mutatja meg az objektumok (esetünkben a szakágazatok) elhelyezkedését. Ez közlési célokra az előbbinél lényegesen jobban használható.

30. ábra. Az élelmiszer-feldolgozás szakágazatainak elhelyezkedése két ismérv szerint



Forrás: Jansik; 2000.

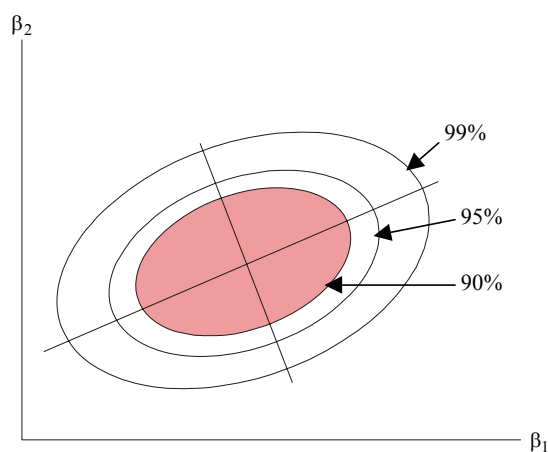
A 30. ábrán a pontdiagramba bejelöltük a mesterségesen képzett osztályokat, így jól látható azok elkülönülése, és az, hogy a részletes szakmai elemzés során lehet-e, és ha igen, akkor hogyan lehet az egy klaszterbe tartozó ágazatokat együtt elemezni, illetőleg egy-egy kiemelt ágazattal a csoportot jellemezni.

Az eddigiekben a leíró statisztika ábráival foglalkoztunk, ám a grafikus ábrázolás a következtetési statisztikában is helyet kap. Nem foglalkozunk külön a becslőfüggvények eloszlásait leíró ábrákkal, ezeket az eloszlások leíró jellegű tárgyalásánál említettük. A teljességre való törekvés igénye nélkül bemutatunk azonban néhány olyan ábrát, amelyek kifejezetten a becslésekhez, illetőleg a hipotézisvizsgálathoz kapcsolódnak, azoknak szerves részét alkotják akár a statisztikus belső eszközöként, akár az eredmények közlésének elősegítőjeként.

A robusztus becslések értékelésekor használatos grafikus eszköz az *érzékenységi görbe* (SC), amely azt mutatja, hogy egy újabb elem bekapcsolása a megfigyelések közé hogyan befolyásolja a különféle becslőfüggvények tulajdonságait. (Lásd például Hunyadi; 2001.) Az intervallumbecslések esetén a konfidencia-intervallumok kiváltképp kétdimenziós esetben jól ábrázolhatók. A regressziószámítás egyik ismert eredménye például az, hogy normális lineáris modellben két magyarázó változó esetén a becslt paraméterek egymástól nem függetlenek, így adott megbízhatósági szint esetén a konfidencia-tartomány a két együttható terében egy kúpszeletet rajzol ki, mely általános esetben a derékszögű tengelyekkel szöget bezáró tengelyű ellipszis.

A 31. ábra ilyen konfidencia intervallumokat mutat különböző megbízhatósági szintek esetén. Az ellipszis tengelyeinek hossza az egyes becslt paraméterek varianciáival arányos, a főtengely vízszintes tengellyel bezárt hajlásszögének koszinusza pedig megegyezik a két becslt paraméter korrelációs együtthatójával. Az ilyen alakú ábrák igen hasznosak a regressziós modellek értékelésénél.

31. ábra. A kétváltozós regressziós modell becslt együtthatóinak konfidenciatartományai

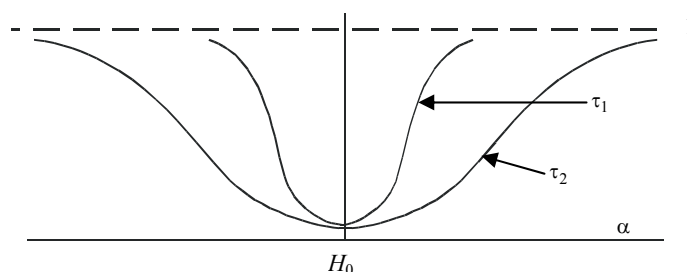


A hipotézisvizsgálatban a becsléseknél elterjedtebb a grafikus ábrázolás alkalmazása. A következőkben három egyszerű példát mutatunk be erre. A tesztek tulajdonságainak értékelésekor a próbafüggvény tulajdonságait leginkább szintetikus módon egy ábra

(pontosabban egymással szoros kapcsolatban álló két ábra) hordozza. Ez a két ábra az OC (Operating Characteristic) görbe és komplementere az *erőfüggvény*. A kettő közül az erőfüggvény az, amelyik gyakrabban használatos: két vagy több teszt tulajdonságainak (torzítatlanság, erő, konzisztencia) összevetésekor az erőfüggvények alapján hozható döntés. Mivel az erőfüggvény az ellenhipotézisek függvényében a hibás nullhipotézis elutasításának valószínűségét adja meg, nagy, 1-hez közeli értékei jelzik valamely rögzített mintanagyság és szignifikancia-szint esetén a jó próbát.

A 32. ábrán két próba erőfüggvényét vázoljuk. A τ_1 -gyel jelölt próba erőfüggvénye az ellenhipotézisek teljes tartományán magasabb értéket vesz fel, mint a τ_2 erőfüggvénye, ezért a τ_1 próbát egyenletesen erősebbnek nevezzük a τ_2 -nél.

32. ábra. Az erőfüggvények ábrázolása

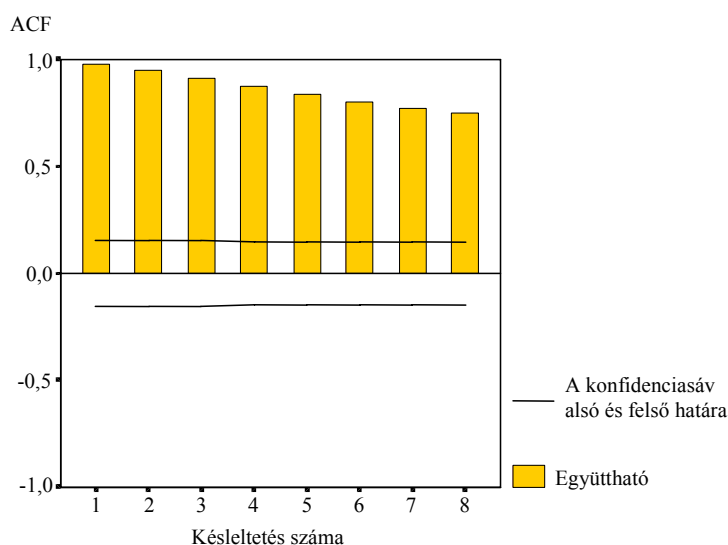


A grafikus ábrázolás külön fejezetét jelentik a grafikus tesztek. Ezúttal tehát nem a tesztek tulajdonságainak értékeléséről, hanem magukról a tesztekéről mint döntési szabályokról beszélünk. A grafikus tesztek egy része a hagyományos tesztek ábrázolt formája; ezek a tesztek valójában nem tekinthetők önálló eszközöknek. Arra minden esetre jók, hogy segítségükkel bizonyos helyzetekben gyors döntéseket hozunk. Egyszerű példa erre az, amikor egy sztochasztikus időszormodell esetében az azonosítás fázisában használt ACF (autokorreláció függvény) alakját és egyes értékeinek szignifikanciáját vizsgáljuk. Maga az ACF is fontos grafikus eszköz a döntés (modellválasztás) szolgálatában, hiszen elsősorban alakja alapján ránézésre azonosítjuk az induló modellt, de a döntést megkönnyítendő a szoftverek grafikus úton, sáv formájában megadják a különböző rendű autokorrelációs együtthatók konfidencia-intervallumait, ami alapján az adott késleltetésű változó jelentősége azonnal (ránézésre) megítélhető.

A 33. ábrában egy ACF-et mutatunk be,⁴ amelynél a szoftver berajzolja a 95 százalékos megbízhatósági sávot. Az ábrából levonható következtetés (döntés) az, hogy a mögötte álló idősor (a Richter Gedeon részvényeinek napi záró árfolyama 1995 első 5 hónapjában) határozott elsőrendű autoregresszív jeleket mutat, így indulásképp egy AR1-gyel (elsőrendű autoregresszív modell) lehet modellezni. Ezt az mutatja, hogy az ACF értéke fokozatosan csökkenő, és valamennyi vizsgált értéke (az első 8 autokorrelációs együttható) 5 százalékon szignifikáns mértékben különbözik 0-tól (az oszlopok kinyúlnak a 95 százalékos megbízhatóságot jelölő sávból).

⁴ Ebben az esetben megemlítjük, hogy a korábban javasoltak értelmében az ACF-et, tekintve, hogy diszkrét értékekre van értelmezve, pálcadiagrammal kellene ábrázolni. A jobb áttekintés kedvéért a számítógépes csomagok gyakran oszlopdiagram formájában ábrázolják az ACF-et. Ez valójában nem helyes, de mivel az ACF többnyire csak a statisztikus belső munkájának eszköze, talán nem zavarja meg a statisztikát kevésbé ismerőket, a statisztikusoknak pedig a kapott eszköz birtokában el kell nézniük ezt a kis pongyolást.

33. ábra. Idősor ACF-ábrája 95 százalékos megbízhatósági sávval



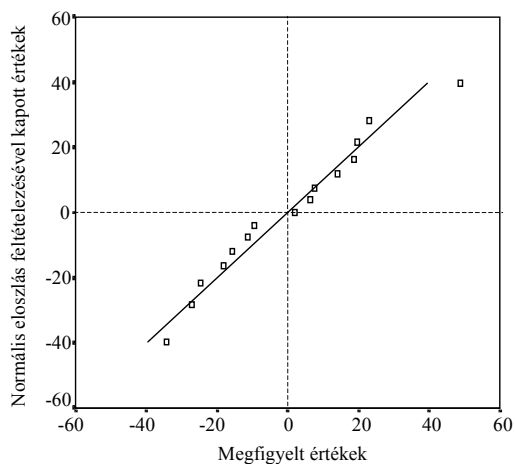
Forrás: Hajdu et al.; 1999.

Vannak azonban olyan grafikus tesztek, amelyek esetén csak az ábra alakja, elrendezése alapján dönthet a statisztikus arról, hogy induló hipotézisét fenntartja-e vagy elveti. Ezen tesztek közül csupán az eloszlásvizsgálatra (és kiemelten a normalitásra) irányuló Q-Q (quantile-quantile) tesztet mutatjuk be. Ennek a tesztnek az alap gondolata az, hogy alapadatainkat standardizáljuk, majd elkészítjük empirikus eloszlásfüggvényüket. Ezt követően az eloszlásfüggvény értékeit a normális eloszlásfüggvény inverze szerint transzformáljuk, és ha az eredeti eloszlás valóban normális volt, akkor a kiinduló adatok és az oda-vissza transzformált értékek megegyeznek, azaz, ha ezek mind-egyikét egy-egy pont két koordinátájának tekintjük, akkor az így kapott pontok egy egyenesen helyezkednek el. Ha a változó csak közelítőleg normális (és természetesen a gyakorlatban ez a jellemző eset), akkor a pontok egy origón átmenő 45 fokos egyenes körül szóródnak. Az így készített pontdiagramot, amely tehát kedvező esetben egy 45 fokos egyenest rajzol ki vagy közelít meg, Q-Q ábrának nevezzük, és a normalitás grafikus tesztelésére használjuk. A tesztelés természetesen nem olyan szigorú, mint a hagyományos tesztek esetén, hiszen a döntés (elfogadás vagy elutasítás) csak ránézés alapján történik, ám az esetek jó részében egy ilyen ábra hozzáértő megtekintése helyes döntést eredményez.

A 34. ábra egy Q-Q tesztet mutat be. Ebben a feladatban egy olyan regressziós becslés maradékait vizsgáltuk, ahol a megyénkénti jövedelemeltéréseket kívántuk magyarázni egyrészt a munkanélküliségi rátával, másrészt az iparban és az építőiparban foglalkoztatottak arányával. A regressziós becslések értékelésének lényeges eleme a modell kiinduló feltételeinek ellenőrzése. A maradékváltozó normális eloszlását Q-Q ábrával teszteljük.

A 34. ábra pontjai jól illeszkednek a megfelelő egyenesre, ezért döntésünk az ábra alapján az, hogy a becslések tükrében a maradékváltozó normális eloszlását nincs okunk kétségbe vonni.

34. ábra. Egy regressziós modell
maradékváltozójának normalitásvizsgálata Q-Q ábrával

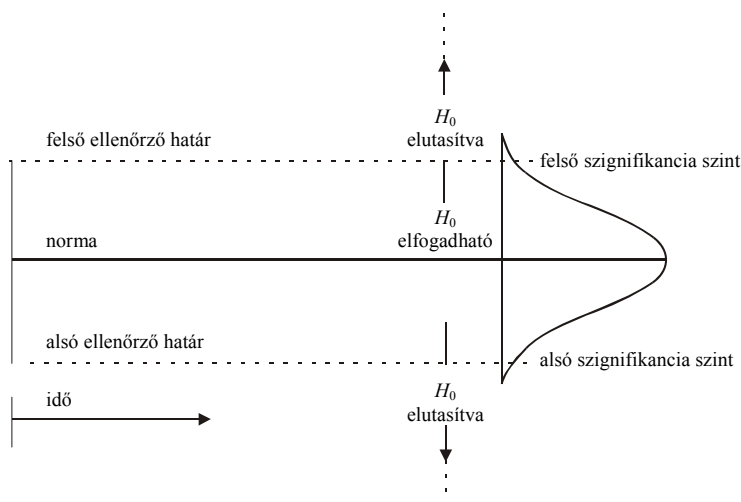


Adatforrás: Területi statisztikai évkönyv, 1997 (1998). Központi Statisztikai Hivatal. Budapest.

Megjegyezzük még egyrészt azt, hogy a Q-Q ábrák mellett ritkábban bár, de használjuk olykor a P-P (probability–probability) típusú ábrákat hasonló céllal, másrészt azt, hogy ezek a tesztek a számítógépek alkalmazása előtt is léteztek; ekkor az ábrázolást az e célra külön készített eszköz (az ún. Gauss-papír) segítette.

Végezetül szólni kell a minőségellenőrzés által használt ún. kontrollkártyákról, amelyek szintén egy grafikus teszt eszközeiként foghatók fel.

35. ábra. A minőségellenőrző kártya szerkezete



Forrás: Statisztika. Emelt A szint; 1990.

A 35. ábra egy kontrollkártya szerkezetét vázolja. Az ábra jobb oldalán a jelenség feltételezett mintavételi eloszlását mutatjuk, amely alapján felrajzolhatók az adott valószínűség melletti tűrési határok. (Például egy átlagos érték tesztelésére normális mintavételi eloszlás feltételezésével könnyen lehet konfidencia-határokat adni.) A kártya ezek után úgy működik, hogy a minőségellenőr (akinek az elméleti háttérrel semmit sem kell tudnia) megfelelő időnként elvégzi az előírt méréseket, és ezek eredményeit felvezeti a kártyára. Ha az eredmények az ellenőrző határokon kívül esnek, ezt minőségi problémák jelzéseként kell felfogni, és ha egy bizonyos időtartományban az ilyen esetek sokszor fordulnak elő, a gyártást felül kell vizsgálni. Megjegyzendő, hogy manapság ezt a feladatot már jórészt folyamatkövető számítógépes rendszerek végzik, de a rendszer működésének elve ugyanaz, mint a bemutatott kártyáé.

A statisztikai munka során még további ábrázolási lehetőségek is adódnak. Elsősorban belső használatra és nem a tájékoztatás igényével egyre több, speciális feladatok megoldását elősegítő munkaábra (célábra) készül. Ezt a folyamatot nagymértékben segíti a számítógépek grafikus eszközei. A statisztikusok munkájuk során különféle blokk-sémákat használnak, egyes mintavételi tervek kialakításakor célábrákon terveznek (görög–latin négyzetek), a különböző folyamatok lefolyását szimulációs ábrákon követik, egyes becslőfüggvények tulajdonságait, a tulajdonságok közti átváltási (trade-off) ábrán (például nyomábra) mutatják be, a döntésekhez döntési fákat használnak stb. Ezeknek még csak felsorolására sem vállalkozunk. Megemlítünk azonban még két lehetőséget, amelyek a hagyományos ábrázolási módokat kibővítik, és egyben új távlatokat nyitnak. Mindkét eszköz a számítógépes ábrázoláshoz kapcsolódik.

Az elsőnek az *interaktivitása* jelent újdonságot. Egyes programcsomagok lehetőséget kínálnak arra, hogy például egy regressziós összefüggésben a képernyőn megjelenítik a megfigyelt pontokat, berajzolják és kiírják a regressziós modell jellemzőit. Ezután a program lehetőséget ad arra, hogy bármelyik megfigyelési pontot a képernyőn módosítsuk, és azonnal lássuk ennek eredményét. Ez az eljárás valójában arra szolgál, hogy a modell lényeges tulajdonságait letapogassuk, megkeressük az érzékeny és a kiugró (outlier) adatokat.

A másik említésre méltó eszköz a *mozgó ábra*. Természetesen ez is csak számítógépes környezetben alkalmazható. Mozgó ábrával lehet a statisztikai folyamatokat kialakulásukban, kifejlődésükben bemutatni. Idősorok lefutását, regressziós összefüggések megbízhatóságát, általában mindennemű szimulációs vizsgálat eredményét kialakulásában lehet szemléltetni, aminek az a hatalmas előnye a statikus ábrákkal szemben, hogy az alkalmazó megtapasztalhatja, milyen törekeny egyetlen (többnyire kis) mintája, illetőleg az abból levonható következtetés (általánosítás) mennyire helytálló.

*

A grafikus ábrázolás a statisztikai munka és kiváltképp az eredményközlés egyik hatékony módszere. A számítógépek elterjedése egyfelől ugrásszerűen növelte az ábrázolás lehetőségeit, másfelől nagyobb odafigyelést és némi szakértelmet követel meg az alkalmazóktól annak érdekében, hogy az ábrázolás helyes és hatékony legyen.

A magyar gazdaság- és társadalomstatisztikai elemzések gyakran élnek a grafikus ábrázolás lehetőségeivel. Az elemzők és felhasználók azonban többnyire csak a legegyszerűbb ábratípusokat használják, azokat sem mindig helyesen. A leggyakrabban elkövetett hibák (melyekre részletesen is felhívja ez az írás a figyelmet): az ábra típusának helytelen vagy félrevezető megválasztása, a zsúfolt, áttekinthetetlen, egyszerre több jelenséget bemutatni kívánó ábrák, a feleslegesen használt eszközök (3D, színek) és a területarányos-

ság megsértése. A hibás vagy nem hatékony ábrák éppen a grafikus ábrázolás lényegét csorbítják, hiszen vagy nem lehet az ábra alapján azonnal és meggyőzően látni azt, amit a szerző sugall, vagy ami még rosszabb, az ábra más következtetést enged meg, mint ami valójában az alaposabb elemzésből adódik.

A helytelen gyakorlat másik forrása az, hogy az elemzők csak a legegyszerűbb, legközönségesebb ábratípusokat használják, holott a statisztika számos szemléletes és a lényegét jól megragadó ábrát fejlesztett ki. E tanulmány utolsó része ezekre mutat be több követendő példát. Új gondolatnak minősül az interaktív és mozgó ábrák kérdése, amelyeket számítógépes környezetben lehet alkalmazni egyes statisztikai jelenségek még hatékonyabb megismerése érdekében.

A vizsgálat egyik fontos megállapítása az, hogy a közkezen forgó programcsomagok ugyan általában sokat segítenek az ábrázolásban, de nem ritka az az eset, amikor félrevezető ábrákat kapunk. Ezek kiigazítása gyakran nem egyszerű. Végezetül megemlíjtük, hogy az itt kifejtettek csak egy – bár sok tapasztalaton alapuló – véleményt tükröznek. Várható, hogy válaszként sok megjegyzés érkezik a helyes és a helytelen alkalmazásokra vonatkozóan. Ezek az írásos hozzászólások sokat segíthetnek abban, hogy kialakítsuk a követendő helyes gyakorlatot.

IRODALOM

- JANSIK, Cs. (2000): Foreign Direct Investment in the Hungarian food sector. *Hungarian Statistical Review*, 78. évf. Special Number 4. 78–104. old.
- JOUBERT K. – GYENIS Gy. (2001): *A 18 éves sorköteles ifjak egészségi állapota, testfejlettsége*. Népeségtudományi Kutatóintézet Kutatási jelentései 70. Budapest.
- HAJDU O. – HUNYADI L. – VITA L. (1999): *Statisztikai elemzések*. (Egyetemi jegyzet.) Budapesti Közgazdaságtudományi és Államigazgatási Egyetem, Budapest.
- HUNYADI L. – MUNDRUCZÓ Gy. – VITA L. (2000): *Statisztika*. (3. kiad.) AULA Kiadó, Budapest.
- HUNYADI L. (2001): *Statisztikai következtetésemélet közgazdászoknak*. Központi Statisztikai Hivatal, Budapest.
- KOTZ, S. – JOHNSON, N. L. (szerk.) (1993): *Encyclopedia of statistical sciences*. Wiley & Sons Inc., New York–Toronto–Brisbane.
- RAPPAL G. (2000): Az uniós csatlakozás időpontja egy üzemgazdasági elemzési módszer tükrében. *Statisztikai Szemle*, 78. évf. 12. sz. 981–993. old.
- Statisztika. Emelt A szint* (1990). Egyetemi jegyzet, Budapesti Közgazdaságtudományi Egyetem, Budapest.

SUMMARY

Some problems of the graphical display are discussed in the paper. First the basic concepts and rules of graphical display are treated. This is followed by a step by step analysis of displaying the commonly used statistical problems. Good and bad practices are demonstrated through a number of examples which cover the wide range of applied statistics. In the concluding paragraph some examples of more sophisticated exercises and special task-forced figures are highlighted.