

BÓDI ANIKÓ

MTA Pszichológiai Intézete, Budapest

MATEMATIKAI STATISZTIKAI EREDMÉNYEK VIZUÁLIS MEGJELENÍTÉSE

Előzmények

Egy kutatás során nyert adatok matematikai eszközökkel való feldolgozása során a kapott eredmények feldolgozását több fázisban végezzük. Az első szakaszban csak az adott matematikai eljárással kapcsolatos ismereteinket, tapasztalatainkat mozgósítjuk: az adatok szerkezetét vizsgáljuk anélkül, hogy szakmai szempontokat érvényesítenénk. A kivitelezéshez papírt és ceruzát ragadunk, apró döntések eredményeképpen egyes számokat bekarikázunk, nyilakat húzunk, sokatmondó jeleket helyezünk el a papír különböző zugaiban. Esetenként kockás vagy milliméter papíron próbáljuk meg a magunk számára szemléletessé tenni a számokat, hogy megtaláljuk azt az elrendezést, amelyet tartalommal töltve szakmai következtetéseket vonhatunk le.

Eleinte élvezettel tölt el minket ez a munka is, idővel azonban szívesen lemondanánk az egyre mechanikusabbá váló műveletek elvégzéséről. Felmerül az igény, hogy ezeket a feladatokat a számítógépre bizzuk, találjunk olyan matematikailag korrekt eljárást, amelynek segítségével a gép maga helyezi el az általunk megszokott jeleket, vagy azok számára kivitelezhető megfelelőit.

Persze amint új lehetőségekkel ismerkedünk meg, igényeink is megnőnek, és már az sem tűnik képtelenségnek, hogy olyan információkat tegyünk szemléletessé a gépi ábrázolással, amire kézzel sohasem vállalkoznánk.

Az utóbbi években a felhasználókkal szoros kapcsolatot fenntartó matematikusok egyre nagyobb hangsúlyt fektetnek arra, hogy az ilyen igényeknek eleget tegyenek. (Jó példa erre a BMDP programcsomag, melynek kétévenként bővített változataiban időről időre nő az ábrázolási lehetőségek száma.) Sajnos a számítógépekhez csatlakozó nyomtatók lehetőségei elég szűkösek az általuk ismert nyomtatható jeltípusok korlátozott száma és szabott méretei miatt.

Az alkalmazott eszköz

Az itt bemutatásra kerülő ábrák egy speciális rajzológép, az úgynevezett plotter lehetőségeinek felhasználásával készültek. Ez a gép ugyanúgy a számítógépen keresztül kapott instrukciókat hajtja végre, mint a nyomtató, de eltérő felépítése miatt felbontása finomabb, utasításkészlete bővebb. Nemcsak nyomtatható jeltípusokat írhatunk vele, hanem tetszőleges görbét is rajzolhatunk, különböző tollakat használva színes ábrát is létrehozhatunk.

Tág lehetőségeit kihasználva (azaz jól tervezett és megírt számítógépes programmal) gyakorlatilag tetszőleges ábrázolásra alkalmas.

A következő eljárásokat megvalósító programok az MTA Pszichológiai Intézetének TPA 1140-es számítógépén íródtak, FORTRAN nyelven, BBC Goerz 281-es plotterre.

A példák

Az eljárásokat a kutatás vezetőjének engedélyével a „Vállalkozói és hivatalnoki magatartás szociálpszichológiai vizsgálata” című kutatás különböző fázisaiban készült ábrák illusztrálják.

A kutatás során különböző csoportoknak a mai magyar gazdasággal kapcsolatos véleményét vizsgálták szociálpszichológiai szempontok szerint. Pszichológiai eszközként Osgood-féle szemantikus differenciált alkalmaztak. Az adatokat többek közt faktoranalízissel elemeztük, a csoportok összehasonlítását varianciaanalízissel végeztük.

A faktormátrix ábrázolása

A faktoranalízis eredményeinek értelmezésekor a legnagyobb figyelem az eredeti változók és a faktorok korrelációinak elemzésére esik. Az ezeket az együtthatókat tartalmazó mátrixot nevezzük faktorstruktúra- vagy röviden faktormátrixnak, p változóra és m ($\leq p$) faktorra e mátrixnak p sora és m oszlopa van. Amennyiben az eredeti változók tartalma jól ismert, a k -adik oszlopban található együtthatók segítségével értelmezhetjük a k -adik faktort. Ugyanakkor a j -edik sort tekintve a j -edik változó legjobb faktorösszetételét kapjuk.

Ezért rendkívül fontos, hogy egyszerre lássuk át valamelyik sort vagy oszlopot. A hagyományos printer-listákon ugyan nagyon pontosak a számértékek, viszonylag sok változó esetén azonban az egyidejű áttekintés nem könnyű feladat.

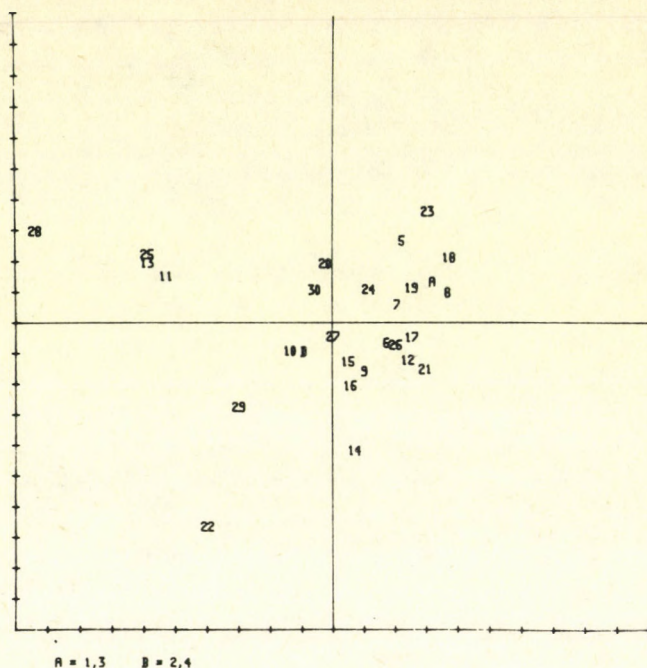
A nagyobb együtthatókat aláhúzással, bekarikázással kiemeljük, majd a még nagyobbakat újra aláhúzzuk, aztán, amikor a lista már kezd áttekinthetetlen lenni, egy újabb papírra kigyűjtjük a fontosnak tűnő információkat.

Ehhez a munkához nyújt segítséget a plotteres ábrázolás. Ekkor az együtthatókat nem számokkal ábrázoljuk, hanem nagyságrendjükkel egyenes arányban telített négyzetekkel. Az együttható előjelét a satírozás irányával, illetve más-más színnel jelöljük.

Így az együtthatók érzékletesebb formában jelennek meg, ráadásul (nem tévedés!) ötvened méretű papíron, ami szintén az áttekintést könnyíti meg.

Egy faktor értelmezésére azt mondjuk, hogy azon változókból tevődik össze leginkább, melyekkel legerősebben korrelál. Hogy melyek ezek a változók, az általában eléggé szembeutó a négyzetek telítettségéből is, de kiemelhetjük ezeket úgy is, hogy a megfelelő négyzetekbe egy kört rajzolunk.

Ha egyetlen lényeges faktort találunk, eligazodunk a számértékeket vizsgálva is. Két lényeges faktort találva a megfelelő faktorértékeket milliméter papíron ábrázoljuk. Az alábbi ábra az ennek megfelelő gépi rajz, amely pontos és elegáns, ráadásul nem egészen 1 perc alatt elkészül.

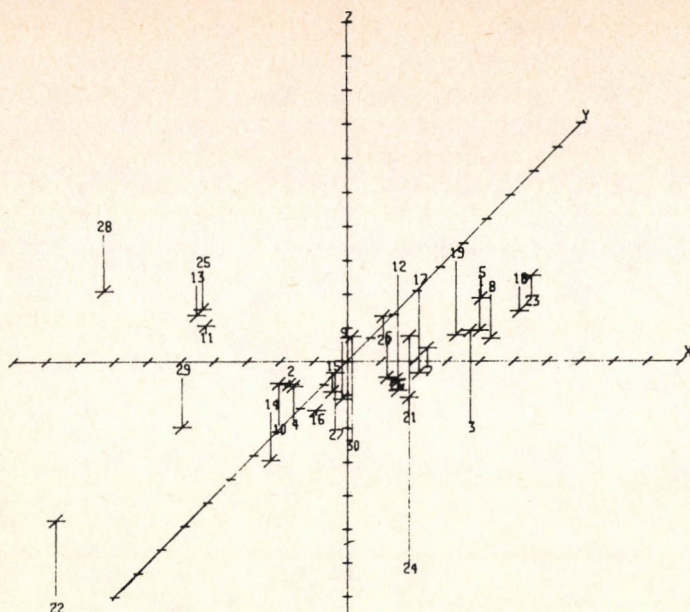


2.1. ábra

Ha a harmadik faktor súlya is nagy, a kétdimenziós ábrázolás félrevezető, ha két egyed az első két faktor szerint hasonló értékeket vesz föl, a harmadikban azonban erőteljesen eltér. A háromdimenziós manuális ábrázolás azt feltételezné, hogy a rajz készítőjében kiváló műszaki rajzoló rejlik.

A géppel ábrázolva a három faktort bal–jobb irányban az első, előre–hátra irányban a második, föl–le irányban a harmadik faktort ábrázoljuk. A rajzon ezek elnevezése rendre X, Y, Z. Ebben a térben úgy helyezzük el az egyedeket, hogy a vízszintes (X,Y) síkon kijelöljük a talppontot, a Z irányú értéket a talpponttól húzott függőleges vonal hossza képviseli. A vonal végpontjában tüntetjük föl az egyed sorszámát.

A következő példában a hivatalnokok háromdimenziós percepciók terében helyezzük el a fogalmakat.



2.2 ábra

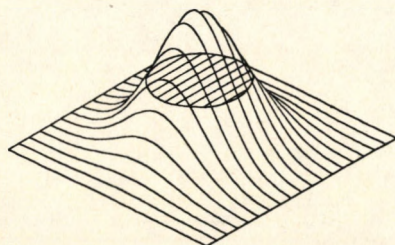
2.2 ábra

Két normális elosztású változó kapcsolatának ábrázolása

Amikor megpróbáljuk elképzelni, hogyan függ össze két változó az általunk vizsgált populációban, a változókat koordináta-tengelyeként kezelve a papírra kisebb-nagyobb, szikárabb-kövérebb ovális alakzatokat rajzolunk. Érzékünk azt sugallja, hogy az alakzat nagysága attól függ, mennyire szóródnak az egyedek változónként külön-külön, kövérsége pedig attól, hogy a két változó mennyire függ össze. Úgy sejtjük, hogy nagyobb szóródáshoz nagyobb alakzat, nagyobb összefüggéshez soványabb tartozik. Nézzük meg, igazolja-e a matematika ezeket a sejtéseinket!

Az adatok kétdimenziós együttes eloszlását vesszük figyelembe. Ezt meghatározza a változók várható értéke és kovariancia mátrixuk (azaz az ebből számítható szórássok és a korrelációs vagy a regressziós együttható). Igazolható, hogy az a görbe, melyen a sűrűségfüggvény azonos értéket vesz föl, éppen egy ellipszis.

Ezt a tényt szemlélteti a következő ábra (amely szintén plotterrel készült, egy olyan általános programmal, amelyet felületek rajzolására írtak).

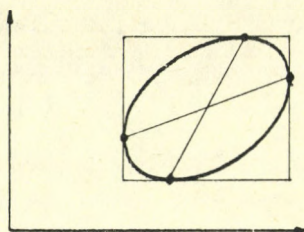


3.1. ábra

Azt kell megvizsgálnunk, hogyan függnek az ellipszis adatai az eloszlás paramétereitől.

Ugyanaz az ellipszis többféleképpen is meghatározható. Megadhatjuk (1) négy pontjával vagy (2) középpontjának, nagy tengelye hajlásszögének, a két tengely arányának és egy, a nagyságára jellemző konstansnak a meghatározásával.

Az első megközelítésből származó információk

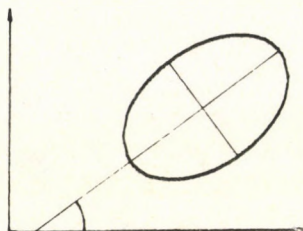


3.2 ábra

Ebben az esetben célszerű a négy pontként az ellipszis „legalsó”, „legfelső”, „legbaloldali” és „legjobboldali” pontját kitűzni. Ekkor az oldalsó pontokat összekötve a függőleges irányban ábrázolt változónak a vízszintesre, az alsó-felső pontokat összekötve a vízszintesnek a függőlegesre vonatkozó regressziós egyenesét kapjuk. (Ez a felismerés némi segítséget nyújt annak a ténynek a megértéséhez, hogy a regressziós egyenesek általában nem esnek egybe. Ezt a geometriai interpretációt követve jól látjuk, hogy teljes függetlenség esetén merőlegesek, és csak teljes összefüggés fennállása mellett azonosak.)

A négy pontban a megfelelő koordináta-tengelyekkel párhuzamosan húzott egyenesek egy téglalapot jelölnek ki. E téglalap oldalainak aránya megegyezik a két változó szórásainak hányadosával, nagyságát a szórások és az a konstans szabja meg, amely arra jellemző, hogy az egyedek hány százaléka kerüljön az ellipszisen belülre. Az ellipszisével azonos középpontja pedig a két változó átlagai által kijelölt pontban van.

A második megközelítésből származó információk



3.3 ábra

A középpont és a nagyság tekintetében nem jutunk új információkhoz.

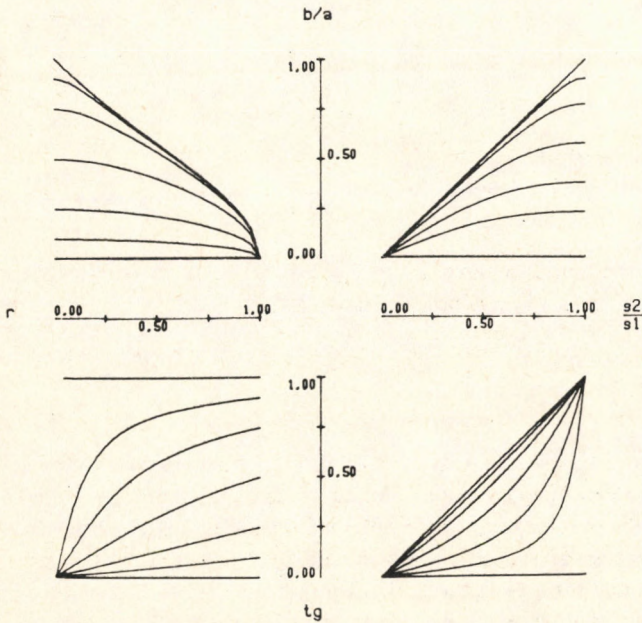
A két tengely arányát és a hajlásszöget a két változó korrelációs együtthatója és szórásaik aránya együttesen szabja meg. A következő ábracsoportból látható, hogy az összefüggés általában nem lineáris.

Az ellipszis kis- és nagy-
tengelyének aránya
a két változó korrelációjának
függvényében,
ha a szórások aránya rögzített.

Az ellipszis kis- és nagy-
tengelyének aránya
a két változó szórásai arányának
függvényében,
ha a korreláció rögzített.

Az ellipszis nagy-tengelyének
hajlásszöge (annak tangense)
a két változó korrelációjának
függvényében,
ha a szórások aránya rögzített.

Az ellipszis nagy-tengelyének
hajlásszöge (annak tangense)
a két változó szórásai arányának
függvényében,
ha a korreláció rögzített.

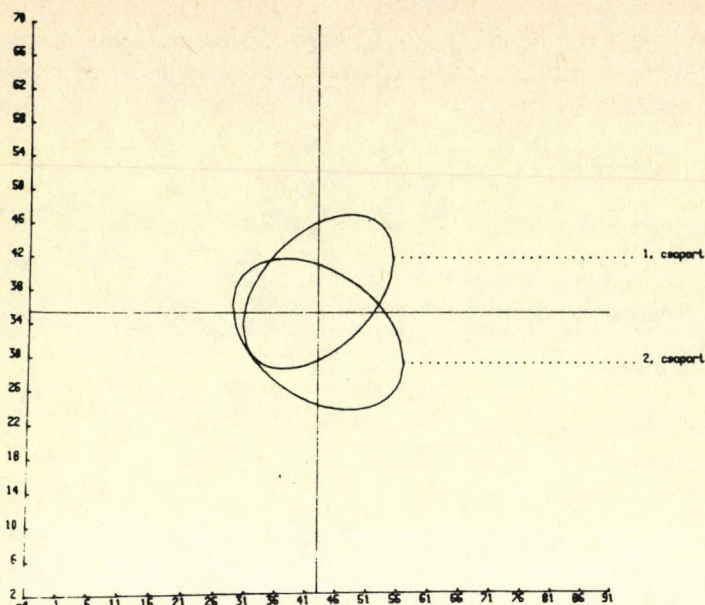


3.4 ábra

A két tengely aránya a szórások hányadosával csak akkor azonos, ha a változók függetlenek. A hajlásszög tangense pedig csak akkor egyezik meg a szórások hányadosával, ha a korrelációs együttható értéke 1. (Ekkor az ellipszis egyenessé fajul.) Érdekes megfigyelni, hogy a szórások egyenlősége mellett a hajlásszög konstans 45° .

A matematika tehát ebben az esetben nagyon hálás partnernek bizonyult: nemcsak sejtéseinket igazolta, hanem olyan egzakt megoldást kínál, amely a vártnál több információt hordoz.

Az ellipszissel való ábrázolás különösen sokatmondó akkor, ha egyazon rajzon több független mintán nyert adatainkat ábrázoljuk.



3.5. ábra

Jelen esetben a „sikeres” és az „igazságos” jelzők összefüggését szemlélteti az ábra a főiskolás (1) és a hivatalnok (2) csoporton.

Összefoglalva

A bemutatott eljárások közösek abban, hogy a számértékeket nem számjegyekkel, hanem tartalmuktól függő vizuális jelekkel ábrázoljuk. Abban azonban eltérnek egymástól, hogy milyen mértékben követtünk kézi eljárást, vagy rugaszkodtunk el attól. Ezt a különbséget éppen a számok tartalma közötti eltérés okozza. Azon közös cél érdekében tehát, hogy jól áttekinthető, figyelmünket a lényeges pontokra irányító segédeszközt kapjunk, egymástól gyökeresen eltérő eljárásokat kell kidolgoznunk.