

## STATISZTIKAI MÓDSZEREK ALKALMAZÁSA A FÖLDTANBAN

BÁRDOSSY GYÖRGY

**Összefoglalás:** A szerző ismerteti a matematikai statisztikának azokat a módszereit, melyeket a földtanban leginkább fel lehet használni. Foglalkozik a statisztikus sokaság gyakorisági eloszlásával és a különböző gyakorisági diagramokkal. Ismerteti az átlagszámokat, valamint a szórást és a ferdeséget kifejező értékmérőket (paramétereket). Végül foglalkozik a kétváltozós korreláció számítással.

A számítási eljárások végrehajtását konkrét példákon mutatja be. A statisztikus adatok földtani felhasználására különböző gyakorlati példákat ismertet.

Napjainkban a természettudományok legkülönbözőbb ágaiban tapasztalható a törekvés az általános megállapítások pontosabb mennyiségi jellegű meghatározására. E célok megvalósítására egyre szélesebb körben alkalmazzák a matematikai statisztika módszereit.

A földtanban és rokon tudományágaiban is egyre inkább térthődítanak ezek a módszerek.

Egyesek azon a véleményen vannak, hogy a földtani folyamatok oly bonyolultak, oly sok tényezőtől tevődik össze egy-egy jelenség és ezek a tényezők oly kevésbé ismertek, hogy hosszadalmas statisztikai számításoknak semmi értelme sincsen: a valóságot úgysem tudják megközelíteni.

Mások maguknak a statisztikus módszereknek földtani létjogosultságát tagadják, merő formalizmusnak tartva az egész eljárást. Számos esetben a statisztikus módszerek hiányos ismerése is hozzájárul ilyen vélemények kialakulásához.

E rossz vélemény alapja részben az, hogy egyesek kellő kritika nélkül nyúltak a statisztika módszereihez, gondolkodást pótló valaminek tekintették, ami minden kérdésre egyszerűen végleges választ ad. Az ilyen mechanikus, minden földtani szemléletet nélkülöző felfogás tévedésekhez vezetett és sajnos gyakran nem a szerzőt, hanem az egyébként jó, de helytelenül alkalmazott módszert diszkreditálta.

E kételyek eloszlatása céljából rövid áttekintésben vizsgáljuk a statisztikai módszerek földtani alkalmazásának lehetőségeit és a módszerek által nyújtott tényleges előnyöket.

A statisztika helyes alkalmazásának első alapfeltétele, hogy a kiértékelésre kerülő adatokat helyes földtani szemlélettel válogassuk össze. Csakis földtanilag összetartozó adatokat értékelhetünk ki közösen és ezen felül még az is fontos, hogy a minták az egész vizsgálandó képződményt egyenletes eloszlásban jellemezzék. Egyes, véletlenül kiragadott minták alapján ugyanis teljesen helytelen eredményekhez juthatunk.

A statisztika segítségével elérhető legegyszerűbb eredmény az adatok rendszerezése, jól áttekinthető beosztása, osztályozása. Gondoljunk például arra, ha egy törmelékes üledékből több ezer mintát vettünk és ezek szemcseeloszlási viszonyait akarjuk összehasonlítani, illetve kiértékelni. Vagy ugyanaz a helyzet, ha egy kőzet

mintáinak több ezer elemzését akarjuk áttekinteni. Ezeket a földtanilag összetartozó adatokat összességükben a statisztika nyelvén statisztikus sokaságnak nevezzük. Míg az egyes adatokat valószínűségi vagy esetlegességi változóknak nevezik.

Az első feladat ilyenkor ennek az áttekinthetetlenül nagy adathalmaznak kisebb csoportokba való összevonása (osztályozása), majd pedig e csoportoknak táblázat, vagy rajz alakban történő jól áttekinthető összefoglalása.

Az osztályok nagyságának helyes kiválasztása a további munka sikerének egyik fő feltétele. Mennél nagyobb egy-egy osztály, annál kevesebb lesz az összes osztály száma — következésképpen annál könnyebben lehet áttekinteni őket. Viszont ugyanakkor az osztályok értéküközének megnövelése egyre inkább elnagyolja az általános képet, egyre kevésbé mutatkoznak meg a finomabb részletek. Az osztályok számának túlzott megnövelése viszont az áttekinthetőséget csökkenti le. Vizsgáljuk meg ezt a kérdést egy konkrét példán, mondjuk egy kőzet vegyi összetételének kiértékelésénél.

Feladatunk a jelen esetben az iszkaszentgyörgyi Kincses-Józsefi bauxitterület vegyi összetételének kiértékelése 2510 vegyelemzés alapján. Tekintettel arra, hogy a bauxit  $Al_2O_3$ ,  $SiO_2$ ,  $Fe_2O_3$  és kötött-víz-tartalmának nagyságrendje általában 1%-nál nagyobb, az elemzések pontossága pedig 0,3—0,5%, legcélszerűbbnek az 1%-os osztályok felvett osztályok látszanak. A  $TiO_2$  esetében ugyanakkor 0,1%-os osztályokat vesszünk fel, tekintettel arra, hogy a másik 4 elemmel szemben a  $TiO_2$  változásai a bauxitban csak tized % nagyságrendűek. Az osztály szűkítésének végső fokon az elemzés pontossága szab határt. A  $TiO_2$  elemzési pontossága 0,1% lévén, értelmetlen lenne a csoportosítást 0,01%-os osztályonként végezni.

Az egyes osztályoknak nem kell szükségszerűen egyforma nagyoknak lenniök. Pl. szemnagysági vizsgálatoknál a kisebb szemnagyságok felé haladva egyre csökkenteni szokták az osztályok értékközeit. Az Atterberg-féle beosztásnál:

2,0	—0,2	mm-en az osztály értékköze	1,8	mm
0,2	—0,02	„ „ „	„	0,18 „
0,02	—0,002	„ „ „	„	0,018 „
0,002	—0,000	„ „ „	„	0,002 „

Az osztály beosztás elkészítése után hozzáfoghatunk az ún. eloszlási vagy gyakorisági táblázat elkészítéséhez. Ez abból áll, hogy megszámláljuk, hogy az összes adat közül hány esik az egyes osztályokba. Pl. fenti példánkon a 2510 vegyelemzés közül a  $SiO_2$ -tartalom esetében a 6—7 százalékos osztályba 173 db elemzés, a 12—13 százalékos osztályba pedig 86 db elemzés esik. Ezeket a számokat nevezzük abszolút gyakorisági számoknak.

A gyakorisági táblázat már ebben a formában is nagyon tanulságos. Leolvashatjuk belőle az értéktartomány alsó és felső szélső értékeit és azt is, hogy melyik osztályban található a legtöbb adat, azaz melyiknek legnagyobb a gyakorisága.

Nehézségbe ütközik viszont, ha ugyanazon képződmény két előfordulási helyét akarjuk összehasonlítani. Ha ugyanis a két előforduláson lényegesen különböző számú adattal rendelkezünk, a gyakorisági számok nagysága annyira különbözni fog, hogy szinte lehetetlen őket összehasonlítóssal kiértékelni. E nehézséget csak úgy küszöbölhetjük ki, ha az abszolút mennyiségeket jelentő gyakorisági számokat relatív mennyiségekké — mégpedig százalékszámokká alakítjuk át. Ezt az átszámítást a rendelkezésre álló adatok összegének arányában végezzük. Pl.: a fent említett bauxitelőfordulás  $SiO_2$  tartalmának 6—7 százalékos osztályában a gyakorisági szám 173 volt. Ennek a számnak 100-szorosát elosztjuk az összes elemzések számával (2510 db) és így 6,89

százalékot kapunk. Ezt a számot viszonylagos gyakorisági százalék-nak nevezzük.

Gyakran fontos azt is tudnunk, hogy bizonyos értékközök felett vagy alatt milyen a vizsgált tulajdonság összesített gyakorisága. Ennek megállapítására szolgálnak az ún. kumulatív gyakorisági százalékok. Kétféleképpen lehet őket megadni. Az egyik esetben vesszük az első osztály gyakorisági százalékát, ehhez hozzáadjuk a következő osztályét. Ez az összeg adja a második osztály kumulatív gyakorisági százalékát. Ezt adjuk hozzá a harmadik osztály gyakorisági százalékához. Az összeadást osztályonként addig folytatjuk, míg az utolsó osztályt el nem értük. Ha az összeadást jól végeztük, akkor ennek hozzáadása után éppen 100 százalékot kell kapnunk. A másik esetben az utolsó osztály felől indulunk ki és haladunk az összegezésével az első felé. A két számsor tagjai ugyanazon az osztályon belül egymást 100 százalékra egészítik ki.

A gyakorlatban általában nem elég a gyakorisági százalékok és a kumulatív gyakorisági százalékok kiszámítása és táblázatos összefoglalása, hanem igyekezni kell ezeket az adatokat rajzban ábrázolni.

Az ábrázolási módok — különböző diagramok stb. — egyik fő célja szemléltetés. Azonban számos esetben a grafikus ábrázolás többet mond a táblázatos vagy szöveges formánál. Olyan körülményekre hívja fel figyelmünket, melyeket különben nem vehetünk volna észre, máskor pedig egyszerű és gyors megoldást szolgáltat hosszadalmas számítások helyett. A grafikus ábrázolástól való idegenkedés egyik jogos alapja, hogy kellő földtani szemlélet hiányában elvonatkoztatja a vizsgált tárgyat annak földtani alapjától. Ez aztán túlzó és így teljesen helytelen következtetésekre vezethet.

A statisztikai adatok grafikus ábrázolásakor azért mindig az adott földtani alapfeltételeket kell szigorúan figyelembe venni és az ábrázolás módját tőlük függően kell kiválasztani.

A gyakorisági adatok ábrázolásakor általában derékszögű koordináta rendszert szoktak használni, melynek vízszintes tengelyére mindig a független változót (osztály), függőleges tengelyére pedig a függő változót (gyakorisági vagy kumulatív gyakorisági százalék) viszik fel.

Kérdéses még a mértékegység kiválasztása. A legtöbb esetben a közismert milliméter beosztást szokták alkalmazni, mely mindkét tengely irányában egyenlő távolságokra osztja fel a papírt.

Használhatnak olyan beosztást is, mely az  $X$  tengely irányában logaritmikus skálát, az  $Y$  tengely irányában pedig milliméteres skálát tartalmaz. Ilyen beosztásra akkor van szükség, ha a vizsgálandó tulajdonság igen tág határok között ingadozik. Jellegzetes példa erre a törmelékes üledékek szemcseceloszlási adatainak ábrázolása. Itt a logaritmikus skála segítségével a kis szemnagysági frakciókat kívánják a nagyobb frakciókkal arányos mértékben ábrázolni.

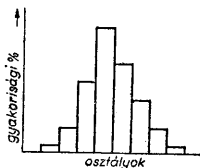
Végül van olyan beosztás is, mely mindkét tengely irányában logaritmikus skálát tartalmaz. Ilyen beosztásra elsősorban akkor van szükség, ha a gyakorisági eloszlás nagyon egyenetlen és ezeket az egyenetlenségeket el akarják simítani.

A logaritmikus skálák alkalmazásakor mindig szem előtt kell tartanunk, hogy ilyenkor tulajdonképpen ugyanazt kapjuk, mintha rendes milliméteres papírra az ill. számok logaritmusait vittük volna fel.

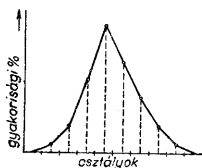
A skálabeosztás kiválasztása után az ún. gyakorisági diagramokat szokták megszerkeszteni. Ezeknek kétféle ábrázolási módja van. Az egyik esetben egy-egy osztályon belül állandónak tekintjük a gyakorisági százalék nagyságát — tehát a gyakorisági százalékokat az osztályok teljes szélességében lépcsőzetesen rakjuk fel (1. ábra).

Az így kapott diagramot gyakorisági hisztogramnak nevezik.

A másik eljárásnál az osztályok középpontjából merőlegeseket emelünk a vízszintes tengelyre és ezekre felmérjük az osztályra vonatkozó gyakorisági százalékot. A kapott végpontokat egyenesekkel kötjük össze és így egy poligont kapunk, melyet gyakorisági poligonnak szoktak nevezni (2. ábra).

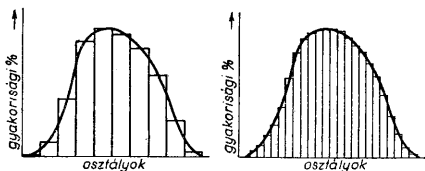


1. ábra. Gyakorisági hisztogram — Frequency histogram — Статистическая гистограмма частотей



2. ábra. Gyakorisági poligon — Frequency polygon — Статистический полигон частотей

A gyakorisági poligon és a gyakorisági hisztogram lényegileg ugyanazt fejezi ki más-más módon. Minél kisebbek az osztályok, annál kevésbé térnek el egymástól. Ha az osztályokat — persze elméletileg — végtelen kicsivé csökkentjük, a két diagram egy ún. gyakorisági görbévé olvad össze. Ez a folytonos görbe kiegyenlíti a két diagram lépcsőszerű ugrásait és sarokpontjait (3. ábra).

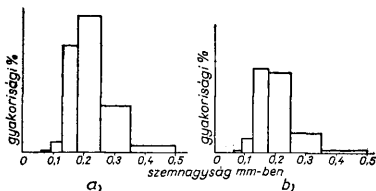


3. ábra. A gyakorisági görbe és hisztogram kapcsolata — Relation between histogram and frequency curve — Взаимоотношение статистической гистограммы частотей и частотей кривой

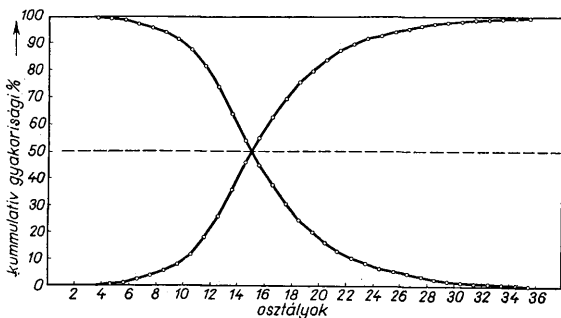
Egyforma nagy értékközönként felvett osztályok esetén ez a szerkesztési mód helyes. Más azonban a helyzet, ha az osztályok értékközei egyre csökkennek, mint pl. szemmagysági frakciók esetén. Ha ilyenkor az előzők szerint rakjuk fel az adatokat, eltorzítjuk a gyakorisági eloszlást a nagy értékközökre vonatkozó osztályok javára. A gyakorisági hisztogramok által határolt területnek ugyanis mindig 100 százaléknak kell lennie az adott mértékegység szerint nézve. Ha viszont az osztályok értékköze nem egyenlő, akkor ez nem adódik ki az első szerkesztési móddal. Ezért ilyen esetekben nem az egyes oszlopok nagyságát, hanem az azok által elfoglalt területet kell alapul vennünk. A 4. ábrán ugyanannak a szemmagysági eloszlásnak a kétféle módon történt ábrázolását láthatjuk. A két hisztogram közt az eltérés bizony eléggé lényeges.

A gyakorisági százalékok ábrázolásával analog módon történik a kumulatív gyakorisági százalékok ábrázolása. Ilyenkor is az  $X$  tengelyre kerül a vizsgált tulajdonság, az  $Y$  tengelyre pedig a kumulatív gyakorisági százalék. Az  $Y$  tengely hossza ekkor a görbe végpontján 100 százalék. A kapott pontokat először poligonszerűen

egyenesekkel kötjük össze, majd pedig ennek segítségével megszerkesztjük a pontokon átmenő legegyszerűbb folytonos görbét. Ezt nevezik kumulatív vagy összeggörbének. Természetesen, amint hogy a kumulatív gyakorisági százalékokat kétféle képpen számíthatjuk, kétféle lehet a görbéjük is. A két görbe egymásnak tükörképi párja és bármely osztálynál 100 százalékra egészíti ki egymást (5. ábra).



4. ábra. Szemcsenagyság eloszlás gyakorisági histogramjai; a) helytelen ábrázolás, b) helyes ábrázolás — Histograms of grain size distribution; a) incorrect method of drawing histogram, b) correct method of drawing histogram — Гистограммы распределения величин зерен: а) неправильный способ построения гистограммы, б) правильный способ построения гистограммы



5. ábra. A gyakorisági eloszlás két összetartozó összeggörbéje — Two coherent cumulative curves of a frequency distribution — Две принадлежащие кумулятивные кривые одной статистической совокупности

A gyakorisági görbe és az összeggörbe szorosan összetartoznak és egymásból kölcsönösen levezethetők. Az összeggörbe nem más, mint a gyakorisági görbe integrálgörbéje és fordítva, a gyakorisági görbe az összeggörbének derivált görbéje. Könnyen megérthetjük ezt, ha arra gondolunk, hogy az összeggörbe minden egyes pontja nem más, mint a gyakorisági görbe százalékainak összege egészen az illető pontig. Azaz a matematika nyelven így írható fel:

$$I = \sum_{i=1}^{i=n} f_i \Delta x$$

ahol

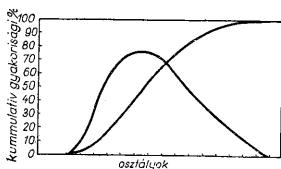
$f_i$  a gyakorisági százalékszámok,  
 $\Delta x$  az osztály értékke,   
 $n$  az osztályok száma a kívánt pontig.

Ha a gyakorisági hisztogramot gyakorisági görbévé alakítjuk át, akkor tulajdonképpen az osztályok értékközeit végtelenül lecsökkentjük. E csökkentés határértékét a matematikában pedig az integrál fejezi ki a következő módon:

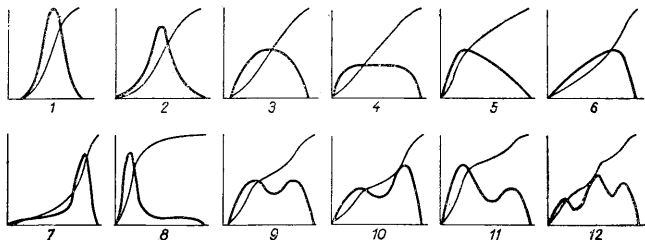
$$\lim_{\substack{\Delta x \rightarrow 0 \\ n \rightarrow \infty}} \sum_{i=1}^{i=n} f_i \Delta x = \int_a^b f dx$$

ahol  $a$  és  $b$  a vizsgált tulajdonság alsó és felső szélső értékei.

Ez az összefüggés nyújt lehetőséget arra, hogy bármely összeggöréből grafikus differenciálás segítségével megszerkeszthessük a hozzátartozó gyakorisági görbét (6. ábra).



6. ábra. Gyakorisági görbe és a hozzátartozó összeggörbe — Frequency curve and the cumulative curve — Статистическая гистограмма частот и принадлежащая к ней кумулятивная кривая.



7. ábra. A földtanban leginkább előforduló gyakorisági és összeggörbék — The most common types of frequency and cumulative curves in geology — Наиболее часто встречающиеся в геологии типы гистограмм и кумулятивных кривых.

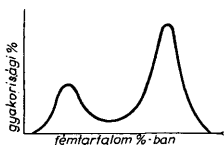
Vizsgáljuk ezután meg, mit nyújtanak a geológus számára a gyakorisági és az összeggörbék?

A 7. ábrán a földtanban leginkább előforduló gyakorisági és összeggörbéket mutatjuk be. Ha a vizsgált tulajdonság — szemmagyság, vegyi összetétel stb. — homogén, úgy szimmetrikus és egymaximumos gyakorisági görbéknek kell kiadódnia (1., 2. sz. görbék). Minél heterogénebb az eloszlás, annál inkább széthúzódik a maximum (3., 4. sz. görbék).

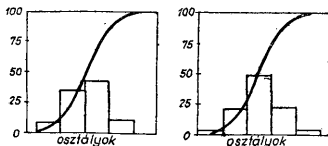
Egyes esetekben a gyakorisági görbe többé vagy kevésbé aszimmetrikus alakú lehet. Ha a leggyakrabban előforduló rész balra tolódik el, balirányú, ha jobbra, jobb-irányú aszimmetriáról beszélünk (5., 6., 7., 8. sz. görbék). Az aszimmetria mindig különleges földtani okok hatására vezethető vissza. Aszimmetrikus vegyi összetétel esetén egyirányban ható vegyi átalakulásra lehet következtetni, mely az eredetileg szimmetrikus gyakorisági eloszlást valamely irányba eltolta. Szemcseösszetéti görbéken akkor

tapasztható leginkább aszimmetria, ha a már leülepedett kőzet anyagát újabb szállító hatás éri, mely belőle bizonyos szemmagyságokat eltávolít. Előállhat ilyen aszimmetria folyami homokra transzgredáló tengeri üledék esetében is.

Végül vannak olyan gyakorisági görbék is, melyeken kettő vagy több maximum figyelhető meg (9., 10., 11., 12. sz. görbék). Az ilyen görbék gyakran igen fontos földtani következtetésekhez szolgáltathatnak alapot. Hogy ezt jobban megérthessük, vizsgáljuk meg az alábbi példát: egy érctelep a földtani megfigyelések alapján egységes felépítésűnek és egyetlen meghatározott folyamat termékének tartanak. Az érc egyik fő komponensének gyakorisági görbéjét megszerkesztették (8. ábra). A többi érces komponensek is hasonló görbéket szolgáltatnak, csak egységnél a nagyobb maximum a jobboldalon, a másikaknál a baloldalon van.



8. ábra. Fém tartalom gyakorisági görbéje egy érc előfordulásán — Frequency curve of a deposit's metal content — Статистическая гистограмма содержания металла одного рудного месторождения.



9. ábra. Egyazon összeggöréből szerkesztett kétféle gyakorisági hisztogram — Two histograms obtained from the same cumulative curve — Две статистические гистограммы, полученные из той же кумулятивной кривой.

Ezek a görbék azt jelzik, hogy az előzetes földtani megállapítások helytelenek voltak és itt tulajdonképpen két, egymásra következő érces fázisról van szó.

Arra, hogy melyik fázisba melyik komponens tartozik, arra a később ismertetendő korrelációs számítás ad választ.

Láthatjuk tehát, hogy a gyakorisági görbék igen lényeges kérdések eldöntéséhez szolgálhatnak alapul.

Az összeggörbék sokkal kevésbé érzékenyek a gyakorisági eloszlás különbségeire, mint a gyakorisági görbék. Értelmezésük ezért nehezekebb és bizonytalanabb. Az összeggörbén az inflexiós pont felel meg a gyakorisági görbe maximumának. Itt a legmeredekebb a görbe.

Az összeggörbéket különösen a szemcseeloszlási viszonyok ábrázolására alkalmazzák nagy előszeretettel. Egyébként mind a gyakorisági, mind az összeggörbékkel egyaránt találkozunk a földtani szakirodalomban.

Az összeggörbék alkalmazásának egyik fő előnye, hogy kevés számú osztály esetén sokkal kevésbé érzékenyek az osztályhatárok felvételére, mint a gyakorisági hisztogramok. A 9. ábrán egyúgyazon összeggöréből szerkesztett kétféle gyakorisági hisztogramot mutatunk be, melyeket úgy nyertünk, hogy az osztályok határait máshol vettük fel. A két görbe eltérése szembetűnő és az értelmezésnél lényeges hibát okozhat. Persze minél több osztályt veszünk fel, annál kisebb az ilyen hiba lehetősége. Tekintettel arra, hogy éppen a szemcseeloszlás vizsgálatakor viszonylag kevés a gyakorlatban használt osztályok száma (5—10), érthető, hogy miért éppen itt alkalmazzák legszívesebben az összeggörbéket.

A görbék értelmezését teljesebbé teszi, ha azok alakjának jellegzetességeit különböző számértékekkel fejezzük ki. A matematikai statisztika a gyakorisági és összeg-

görbék jellemzésére legkülönbözőbb számértékeket dolgozott ki, melyek közül többet a földtani vizsgálódásoknál is sikerrel lehet alkalmazni. Ezeket eloszlás jellemzőknek vagy paramétereknek szokták nevezni.

A paraméterek első csoportja azt fejezi ki, hogy a gyakorisági görbével jellemzett statisztikus sokaság tagjai milyen középérték köré csoportosulnak. Összefoglaló néven átlagszámoknak nevezik őket.

Legismertebb közülük a számtani közép ( $M_{sz}$ ). Ezt úgy kapjuk meg, hogy a statisztikus sokaság tagjainak összegét elosztjuk a tagok számával. Ezt a műveletet a következő képlet fejezi ki:

$$M_{sz} = \frac{\sum_{a=1}^{a=n} a}{n} \quad \text{ahol } n = \text{a tagok száma,} \\ a = \text{a vizsgált tulajdonság nagysága az egyes tagoknál.}$$

Ha a tagokat előzőleg már eloszlási táblázatba vontuk össze, előnyösebb az úgynevezett súlyozott átlagszámítással meghatározni a számtani középértéket. Ilyenkor az egyes osztályok gyakoriságának arányában „súlyozzuk” a vizsgált tulajdonságot. Ennek az eljárásnak általános képlete a következő:

$$M_{sz} = \frac{\sum_1^n a \cdot w}{\sum_1^n w} \quad \text{ahol } w \text{ a vizsgált tulajdonság gyakorisága.}$$

A számítás könnyebb megértése céljából tekintsük az alábbi példát: Egy kőzet mésztartalmának gyakorisági eloszlása 200 kőzetelemzés alapján a következő:

14,0—15,0%	mésztartalom	50%	gyakorisággal
15,0—16,0%	„	25%	„
16,0—17,0%	„	25%	„

A számtani közép meghatározásához ilyenkor nem kell mind a 200 elemzési adatot végigszámolni, hanem a fenti képletet használjuk:

$$M_{sz} = \frac{\sum_1^n a \cdot w}{\sum_1^n w} = \frac{14 \cdot 5.50 + 15 \cdot 5.25 + 16 \cdot 5.25}{50 + 25 + 25} = 15.2\%$$

A súlyozott átlagszámításnak különösen az ásványi nyersanyagkészletek kiszámításánál van komoly szerepe, ahol segítségével igen le lehet egyszerűsíteni az egyébként is hosszadalmas számításokat.

A földtani gyakorlatban átlag alatt szinte kizárólag a számtani közép szerepe van, holott a többi átlagszám nem kevésbé fontos.

A geometriai közép ( $M_g$ ) kiszámítása oly módon történik, hogy a tagokat összeszorozzuk egymással és szorzatukból a tagok számának megfelelő gyököt vonunk.

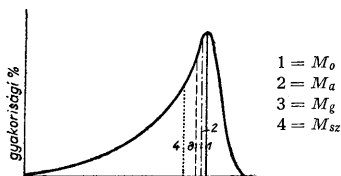
Képlet formájában:  $M_g = \sqrt[n]{a_1 \cdot a_2 \cdot \dots \cdot a_n}$ . E számítás ebben a formában rendkívül fáradságos és hosszadalmas. Meggyorsíthatjuk, ha a gyakorisági százalékok helyett azok logaritmusaival dolgozunk. A számítás ilyenkor leegyszerűsödik a számtani közép képletére. Végül a kapott eredményt logaritmustáblázat segítségével visszakkeressük. A számítás sajnos még így is elég hosszadalmas, ezért a földtani gyakorlatban az átlagszámok közül ezt alkalmazták eddig legritkábban.



A számtani és geometriai közép jellegzetessége, hogy nagyságukat az egész vizsgált sokaság minden egyes tagja befolyásolja. Különösen akkor befolyásolja nagyon, ha a gyakorisági eloszlás erősen aszimmetrikus, vagy kiütő értékek tartkítják. A kiütő értékekre egyébként a geometriai közép kevésbé érzékeny, mint a számtani közép.

A következő átlagszám a medián. ( $M_d$ ). Kiszámítását legkönnyebben úgy érthetjük meg, ha az előzőekben ismertetett kumulatív gyakorisági százalékokra gondolunk. Ott, ahol ezek éppen 50 százalékot érnek el, annál az értéknél van a medián. A medián tehát a statisztikus sokaság számszerint középső tagja. A gyakorisági görbe által bezárt területet a medián két egyenlő nagyságú részre osztja.

Még egy átlagszám van, melyet földtani vizsgálatoknál használni szoktak és ez a módus ( $M_o$ ). Ez azt az értéket jelöli, mely a vizsgált sokaságban a leggyakrabban fordul elő. Az eloszlási táblázatban a módus a legnagyobb gyakorisági százalékkal szerezplő osztály közepén jelölhető ki. A módus földtani szempontból különösen fontos,



10. ábra. A gyakorisági görbe és az átlagszámok kapcsolata — Relation of the several averages and the frequency curve — Отношение различных средних к кривой распределения частот

hiszen a vizsgált adatok legjellemzőbb, típusos értékét jelzi. Sőt az eddig tapasztaltak azt mutatják, hogy a legtöbb esetben éppen a módust tekinthetjük a kérdéses tulajdonság normális nagyságának, a többit csak ettől való eltéréseknek.

A medián és a módus nagyságát a számtani és geometriai középtől eltérően nem befolyásolja a gyakorisági eloszlás esetleges aszimmetriája, vagy kiütő értékei. Előnyük, hogy a meghatározásukhoz nem kell hosszadalmas számításokat végezni, hanem a gyakorisági, ill. az összeggörbéből közvetlenül leolvashatók. A módus a gyakorisági görbe csúcspontján, az összeggörbének pedig inflexiós pontján fekszik. A medián meghatározása csak az összeggörbe alapján végezhető el, mégpedig úgy, hogy megkeressük a görbén az  $Y$  tengely 50 százalékos értékének megfelelő  $X$  értékét.

Ha a gyakorisági eloszlás szimmetrikus, elméletileg mind a négy átlagszám ugyanazt az értéket veszi fel. A természetben az ilyesmi igen ritka, mert minden földtani folyamatot több-kevesebb aszimmetria jellemez. Aszimmetrikus eloszlás esetén a 4 átlagszám széthúzódik egymástól. Mégpedig úgy, hogy a görbe csúcspontján marad a módus. Ezután az elnyúlt ág irányában következnek először a medián, majd a geometriai közép és végül a számtani közép (10. ábra).

Az átlagszámoknak a gyakorisági görbe aszimmetriájától függő jellegzetes elhelyezkedése átvezet a statisztikus mutatószámok (paraméterek) második fontos csoportjához, melyek feladata éppen a gyakorisági eloszlás aszimmetriájának számszerű kifejezése.

Az aszimmetria nagyságát legegyszerűbben úgy fejezhetjük ki, hogy kiszámítjuk a módus és a számtani közép különbségét. Mennél nagyobb a két átlagszám eltérése, annál nagyobb az eloszlás aszimmetriája:  $F = M_o - M_{sz}$ .

Ahhoz, hogy az aszimmetriát kifejező többi mutatót megérthessük, előbb meg kell ismerkednünk a *quartilisek* fogalmával. Ezek a mediánal rokon számok a statisztikus sokaság első és harmadik negyedének határát jelzik. Kijelölésüket szintén az összeggörbe alapján lehet a legegyszerűbben elvégezni. Ott található, ahol az összeggörbe 25 százalékot, illetve 75 százalékot ér el. Jelzésük  $Q_1$ , ill.  $Q_3$ .

Az úgynevezett *számtani ferdeséget* úgy kapjuk meg, hogy a két quartilis számtani közepét összevetjük a medián nagyságával. Vagyis:  $F_{sz} = M_d -$

$$\frac{(Q_1 + Q_3)}{2}$$

Hasonló ehhez a *mértani ferdeség*, melyet a következő képlet szerint

$$\text{lehet kiszámítani: } F_g = \sqrt{\frac{Q_1 \cdot Q_3}{M_d^2}}.$$

A mértani közép kiszámításához hasonlóan a számítási munkát itt is le lehet egyszerűsíteni logaritmusok használatával, amikor is a képlet a következő alakra egyszerűsödik le:

$$\log F_g = \frac{1}{2} (\log Q_1 + \log Q_3 - 2 \log M_d).$$

Az elsőnek említett különbség, valamint a számtani ferdeség közös hátránya, hogy a kapott eredmény függ a vizsgált adatok abszolút nagyságától. Tehát ugyanolyan eloszlást feltételezve, egy kavics „számtani ferdesége” nagyobb lesz egy homokénál. Ezek a mutatók tehát csak azonos nagyságú adatok (pl. azonos szemmagyság) összehasonlítására alkalmasak. Ezzel szemben a geometriai ferdeség az adatok nagyságától független.

Az eloszlásjellemzők (paraméterek) harmadik csoportja azt fejezi ki, hogy a vizsgált adatok milyen mértékben szóródnak szét az átlagértékek körül. Összefoglaló néven *szórásnak* szokták nevezni őket.

A szórás ismerete földtani szempontból igen lényeges lehet, hiszen ez mondja meg, mennyire homogén vagy heterogén a vizsgált tulajdonság szempontjából az illető képződmény. Törmelékeny üledékeknel a szórás számszerű kifejezője az osztályozottság mértékének.

A szórás nagyságának gyors, közelítő pontosságú meghatározására szolgálunk az ún. quartális eltérések.

A számtani quartális eltérést a következő képlet segítségével számíthatjuk ki:  $DQ_{sz} = \frac{(Q_3 - Q_1)}{2}$ .

A mértani quartális eltérés képlete pedig:  $DQ_g = \sqrt{\frac{Q_3}{Q_1}}$ .

Mindkét számérték viszonylag könnyen és gyorsan számítható ki a quartilisek ismeretében. A számtani quartális eltérés hátránya, hogy függ a vizsgált tulajdonság abszolút nagyságától (pl. szemmagyság). A mértani quartális eltérés ezzel szemben független ettől. Szemmagysági vizsgálatoknál ezért az utóbbi használata helyesebb. *Trask* [8] a mértani quartális eltérést *osztályozottsági együtthatónak* nevezte el. Több száz minta vizsgálata alapján megállapította, hogy jól osztályozott törmelékeny üledékek osztályozottsági együtthatója 2,5-nél kisebb, a normálisan osztályozottaké 3,0 körül van, a rosszul osztályozottaké pedig 4,5-nél nagyobb.

Mind a számtani, mind a mértani quartális eltérés csak közelítően fejezi ki a tényleges szórást, hiszen nem vonatkoznak a gyakorisági eloszlás 25 százalékánál kisebb

és 75 százaléknál nagyobb, szélső helyzetű értékeire. A teljes statisztikus sokaság szórását a négyzetes középeltérés (standard deviation) és az átlagos eltérésnégyzet (diszperzió) fejezi ki. Bizonyos matematikai megfontolások miatt, melyekkel itt nem kívánunk részletesen foglalkozni, e számértékeknél az átlagtól való eltérések négyzetének összegét tekintik összehasonlító alapul. Az eltéréseket a 4 átlagszám közül a számtani középtől számolják. Az elméleti számítások ugyanis azt mutatják, hogy az eltérések négyzetének összege akkor a legkisebb, ha azokat a számtani középtől számoljuk.

Az átlagos eltérésnégyzet (diszperzió) nem más, mint az eltérések négyzetének átlaga. A matematika nyelvén ezt a következő képlet fejezi ki:

$$\sigma^2 = \frac{\sum_1^n (x_i - x_0)^2 w}{\sum_1^n w}$$

ahol  $n$  a sokaság tagjainak száma,  
 $x_i$  az egyes osztályok nagysága,  
 $x_0$  a számtani közép nagysága,  
 $w$  az egyes osztályok gyakorisági százaléka.

A négyzetes középeltérés (standard deviation) a diszperzió négyzetgyökével egyenlő:

$$\sigma = \pm \sqrt{\sigma^2} = \pm \sqrt{\frac{\sum_1^n (x_i - x_0)^2 w}{\sum_1^n w}}$$

Mennél kisebb a négyzetes középeltérés nagysága, annál homogénebb a képződmény a vizsgált tulajdonság szempontjából. Törmelékes üledékek esetén annál nagyobb fokú az osztályozottság. A diszperzió és a négyzetes középeltérés végeredményképpen ugyanazt fejezi ki. A diszperzió részletesebb, kis különbségeket jobban kifejez. A négyzetes középeltérés előnye viszont az, hogy a vizsgált tulajdonság nagyságrendjében adja meg a szórást.

Ha a gyakorisági eloszlás nagyjából szimmetrikus, akkor az összes gyakoriságok kb. 2/3-a (68 százaléka) sorol a négyzetes középeltérés által megadott két határérték közé.

A diszperzió és négyzetes középeltérés számításának módját az 1. táblázaton mutatjuk be. A feladat ebben az esetben az iszka: zongyörgyi „Kincses-József” nevű bauxitelfordulás  $\text{SiO}_2$  tartalmának kiértékelése. A táblázat 1—5. sz. oszlopai a gyakorisági százalékok és a különböző átlagszámok a már ismert módon való kiszámítását szolgálják. A 6., 7. és 8. számú oszlopok szolgálnak a szórás kiszámítására. A 6. sz. oszlopba beírjuk minden egyes osztálynak a számtani középtől való eltérését. Ilyenkor az osztályok értékközénel: közepével számolunk, tehát a 10—11 százalékos osztály esetén 10,5-tel. Ezután a kapott eltéréseket négyzetre emeljük és beírjuk a 7. sz. oszlopba. A következő lépés az, hogy minden egyes eltérés-négyzetet összeszorozunk a hozzátartozó gyakorisági százalékkal. A kapott szorzatokat a 8. sz. oszlopba írjuk fel. Végül összegezzük ennek az oszlopnak az adatait és az összeget a gyakorisági százalékok összegével, tehát százal elosztjuk. Az így kapott szám a diszperzió. Ez a jelen esetben 49,22. Ha ebből négyzetgyököt vonunk, megkapjuk a négyzetes középeltérést,  $\pm 7,02$  százalékat. Ez azt jelenti, hogy ezen a bauxitelforduláson a  $\text{SiO}_2$  tartalom számtani középtől az egyes minták  $\text{SiO}_2$  tartalma átlagosan  $\pm 7,02$  százalékkal tér el.

Az átlagot, az aszimmetriát és a szórást kifejező számértékek általában egy-egy képződményen belül állandók és jellemzők rá (kivéve, ha fációs változás lép fel). E mutatószámok (paraméterek) ezért kiválóan alkalmasak földtani korrelálásra, telepazonosításra olyan esetekben, ha ősmaradványok és jellegzetes üledékjellegek hiánya miatt

egyébként erre nem volna mód. Ahhoz, hogy egy ilyen munka megbízható legyen, természetesen nagyobb számú — lehetőleg több száz — minta feldolgozására van szükség.

Felhasználhatók a paraméterek a tektonikában is nagyobb számú rétegdőlés vagy litoklázis adat kiértékelésekor. A gyakorisági görbe ilyenkor megadja a fő tektonikai irányokat, a paraméterek alapján pedig a mozgások jellegére következtethetünk. Különösen tektonikai adatok statisztikus kiértékelése esetén kell az adatok földtanilag helyes összeválogatására nagy gondot fordítani.

Az előzőekben ismertetett paramétereken kívül geológusok és petrográfusok több speciális, egyénileg kiötlött mutatószámot is megpróbáltak a gyakorlatba bevezetni. A legtöbb esetben azonban a szerzők nem vizsgálták meg mutatószámaik elméleti statisztikai jelentését. Ezért aztán az ilyen adatok földtani értelmezése is nélkülözi a szükséges elméleti megalapozottságot. Legismertebb közülük a „B a k e r-féle „szem-nagyság-faktor”, a Niggli és Zringg által kidolgozott „maximum-minimum” módszer és a Purdy-féle „felület-faktor”. Mindhárom módszer törmelékes kőzetek szemmagysági viszonyainak jellemzésére szolgál. E módszerek hazai alkalmazását mászúnkról nem látjuk célszerűnek elsősorban statisztikai-elméleti megalapozatlanságuk miatt.

A földtanban gyakran van szükség valamely képződmény különböző tulajdonságainak összehasonlítására. Például szeretnénk megismerni, van-e valamiféle összefüggés egy törmelékes üledék szemmagysága és szemcséinek koptatottsága között? Vagy egy kőzet mérszartalmát agyagtartalmával akarjuk összehasonlítani.

Ezeknek az összefüggéseknek meghatározásához, sőt az összefüggés mértékének számszerű kifejezéséhez nyújt lehetőséget a matematikai statisztika egyik módszere, az úgynevezett korreláció számítás. Korreláció számítást a természet-tudományok földtannal rokon ágaiban már egyre töbket alkalmazzzák, így pl. a hidrogeológiában is. 1952-ben jelent meg Bogárdi J. „Korrelációs számítás és alkalmazása a hidrológiában” című könyve, melyben nagy matematikai felkészültséggel részletesen ismerteti a különböző korreláció számítási eljárásokat [1].

E dolgozat szűkreszabott keretei közt nem foglalkozhatunk részletesen a korreláció számítás matematikai alapjaival. Itt csupán a korreláció számítás legegyszerűbb formájának, a kétváltozós korreláció számításnak egyik eljárását ismertetjük. Ez az eljárás gyakorlati tapasztalataink szerint leginkább alkalmas a földtani összefüggések számszerű meghatározására és egyaránt alkalmazható bármely tulajdonság, bármilyen más fizikai tulajdonsággal (térfogatsúly stb.) vagy egyéb adattal (mélység, vastagság stb.) való összehasonlítására.

A számítás kiinduló alapja az ún. kétváltozós korrelációs táblázat. Példaképpen bemutatjuk az izamajori bauxitelfordulás  $Al_2O_3$  és  $SiO_2$  tartalmának korrelációs táblázatát (II. táblázat). A táblázat megszerkesztése úgy történik, hogy sorraversezzük a kiértékelendő 1589 db vegyelemzést. Megkeressük az elemzés  $Al_2O_3$  tartalmának megfelelő vízszintes rovatot. Ezután megkeressük az elemzés  $SiO_2$  tartalmának megfelelő függőleges oszlopot. Ahol a két oszlop metszi egymást, vonallal vagy ponttal jelet teszünk. Hasonlóképpen járunk el az összes elemzésnél, majd pedig megszámloljuk, hogy az egyes kockákba hány vonal (elemzés) került és ezt a számot beírjuk az illető kockába.

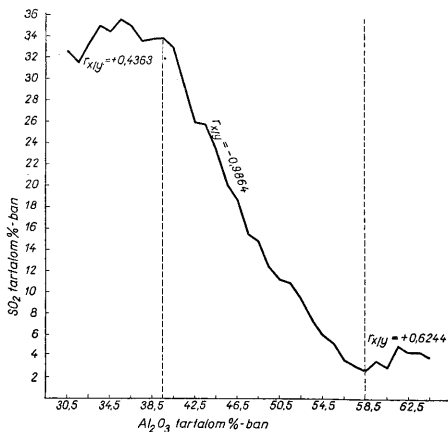
A korrelációs táblázat már ebben a kezdeti formájában is igen tanulságos, hiszen a számok eloszlásának módjából következtetni lehet, van-e összefüggés a két komponens között? Így a példaként közölt II. sz. táblázatból is rögtön láthatjuk, hogy az  $Al_2O_3$  és a  $SiO_2$  tartalom között szoros összefüggés áll fenn. Mégpedig minél nagyobb az  $Al_2O_3$  tartalom, annál kisebb a  $SiO_2$  tartalom.



Ez azonban még nem elég. Tudni akarjuk, hogy százalékonként minden egyes  $\text{Al}_2\text{O}_3$  tartalomnak átlagosan milyen nagy  $\text{SiO}_2$  tartalom felel meg. Ezt is könnyen meghatározhatjuk, ha kiszámítjuk minden egyes vízszintes rovat súlyozott számtani közepét. A súlyozott számtani közép kiszámítását az előzőekben már ismertettük, képlete

$$M_{sz} = \frac{\sum_1^n a \cdot w}{\sum_1^n w}. \text{ A jelen esetben „}w\text{” az illető kockába eső elemzések száma, „}a\text{” pedig}$$

az illető kockához tartozó  $\text{SiO}_2$  érték nagysága. A táblázat jobboldalán a számolás megkönnyítése céljából külön rovatban tüntettük fel a  $\sum_1^n w$  és a  $\sum_1^n wa$  értékeket, végül



11. ábra. Egy bauxitelfordulás  $\text{Al}_2\text{O}_3$  és  $\text{SiO}_2$  tartalmának korrelációs diagramja  $r_{x/y}$  = korrelációs tényező — Correlational diagram of the  $\text{Al}_2\text{O}_3$  and  $\text{SiO}_2$  content of a bauxite deposit.  $r_{x/y}$  = correlational coefficient — Корреляционная диаграмма содержания  $\text{Al}_2\text{O}_3$  и  $\text{SiO}_2$  одного бокситового месторождения,  $r_{x/y}$  = коэффициент корреляции.

pedig belőlük nyert súlyozott számtani közepeket. Ezek azt mutatják meg, hogy pl. 41,5 százalék (azaz 41—42 százalék)  $\text{Al}_2\text{O}_3$  tartalom esetén a  $\text{SiO}_2$  tartalom átlagosan 29,5 százalék: 55,5 százalék  $\text{Al}_2\text{O}_3$  tartalom esetén pedig átlagosan 5,3 százalék stb.

Ha a  $\text{SiO}_2$  tartalomhoz viszonyítva akarjuk az átlagos  $\text{Al}_2\text{O}_3$  tartalmakat megismerni, akkor a fentivel teljesen analóg módon függőleges oszlopok mentén számítjuk ki a súlyozott számtani közepeket. A táblázat adatait még szemléltetőbbé, jobban áttekinthetővé teszi, ha diagram formájában értékeljük ki. A diagramot derékszögű koordinátarendszerben célszerű megszerkeszteni, ahol az X tengelyre az egyik, az Y tengelyre a másik vizsgált tulajdonság értékeit rakjuk fel százalékonként. Jelen esetben az X tengelyre az  $\text{Al}_2\text{O}_3$  tartalom, az Y tengelyre a  $\text{SiO}_2$  tartalom került. Ezután a koordináta tengelyek alapján felrajzoljuk minden egyes  $\text{Al}_2\text{O}_3$ -tartalomnak megfelelő  $\text{SiO}_2$ -tartalmat. E pontokat egymással összekötve készen áll az  $\text{Al}_2\text{O}_3$ - és  $\text{SiO}_2$ -tartalom korrelációs diagramja (11. ábra).

A diagram segítségével pontosan megállapíthatjuk a két komponens összefüggésének jellegét és az összefüggés mértékét.

A fenti példa alapján egészen nyilvánvaló a korreláció számítás jelentősége a földtani folyamatok megítélésében.

A kapcsolat felismerése után még azt a kérdést is tisztáznunk kell, hogy milyen szoros ez a kapcsolat?

A legszorosabb, ideálisan teljes kapcsolat az ún. „függvénykapcsolat”. Ilyenkor az egyik tulajdonság értékének változását a másik tulajdonság teljesen azonos mérvű változása követi. A változás jellege természetesen lehet akár egyirányú (növekedésre növekedés) vagy ellentétes irányú (növekedésre csökkenés).

A természetben ez az eset mondhatni sohasem fordul elő. Szélső értékek tekintetjűk csupán, melyet többé-kevésbé közelítenek meg a kapcsolatok. A másik szélső érték — a kapcsolat teljes hiánya — viszont természetesen bárhol előfordulhat.

E két szélső érték között találjuk az ún. „valószínűség-elméleti kapcsolatokat”. Ezek jellemzik a földtani összefüggéseket is. Ilyenkor a független változó (egyik tulajdonság) értékéhez a függő változó (másik tulajdonság) mindig több, határozott gyakoriságokkal bekövetkező értéke tartozik.

A kapcsolat szorosságát ilyenkor az ún. korrelációs tényező adja meg. A korrelációs tényezőt az alábbi képlet segítségével határozhatjuk meg:

$$r_{x/y} = \frac{\mu_{x/y}}{\sigma_x \cdot \sigma_y}$$

ahol  $\sigma_x$  az egyik tulajdonság négyzetes középeltérése,  
 $\sigma_y$  a másik tulajdonság négyzetes középeltérése,  
 $\mu_{x/y}$  centrális momentum.

A kapcsolat teljes hiánya esetén  $r_{x/y}$  értéke = 0; függvénykapcsolat, tehát a lehetséges legszorosabb kapcsolat esetén pedig  $r_{x/y}$  = 1. A valószínűség-elméleti kapcsolatoknál  $r_{x/y}$  e két szélső érték közt váltakozik. A megadott határértékeken belül a korrelációs tényező lehet pozitív és negatív előjelű. Ha pozitív előjelű, akkor az egyik tulajdonság növekedése a másik tulajdonság növekedését vonja maga után. Ha negatív, akkor a két tulajdonság nagysága ellentétesen változik.

A korrelációs tényező kiszámítását az izamajori bauxitelfordulás  $Al_2O_3$ - és  $SiO_2$ -tartalmának előzőkben említett példáján mutatjuk be. A korrelációs táblázat (II. táblázat) alapján először is megszerkesztjük a korrelációs diagramot (11. ábra). Az ábrából kiderül, hogy az  $Al_2O_3$  és  $SiO_2$  kapcsolata nem mindenütt egyforma. Kis  $Al_2O_3$ -tartalomnál a görbe közel vízszintesen fut, és hasonló a helyzet a nagy  $Al_2O_3$ -tartalomnál is. Közepes  $Al_2O_3$ -tartalomnál viszont a görbe erősen lejt, vagyis azt jelzi, hogy ebben a szakaszban az  $Al_2O_3$ -tartalom növekedésével egyidejűleg a  $SiO_2$ -tartalom erősen csökken. A két komponens kapcsolata tehát a három szakaszban más és más jellegű. A korrelációs tényezőt ezért mindhárom szakaszra külön-külön kell kiszámítani. A III. táblázaton a középső szakaszra (39,5—58,5%  $Al_2O_3$ -tartalom) vonatkozó korrelációs tényező számítását mutatjuk be. A táblázat első oszlopa az  $Al_2O_3$ -tartalom egymásra következő osztályainak középértékeit tartalmazza (50,0—51,0% esetén 50,5 stb.). A második oszlopa a megfelelő átlagos  $SiO_2$  értékek kerülnek. Ezeket a korrelációs táblázat (II. táblázat) megfelelő oszlopából olvassuk le. Ezután kiszámítjuk mind az első, mind a második oszlop számtani középértékeit. A harmadik oszlopba az  $Al_2O_3$  értékeknek a számtani középértéktől való eltéréseit írjuk be, a negyedik oszlopba pedig a  $SiO_2$ -ét. Az ötödik és hatodik oszlopokba az eltérések négyzetre emelt értékeit írjuk be. Ezeket oszloponként összegezzük, az összegeket pedig elosztjuk a tagok számával —

a jelen esetben hússzal. Ezekből négyzetgyököt vonva megkapjuk a két komponens négyzetes középeltéréseit — melyekről az előzőkben már részletesebben szözlünk. Ezután már csak a centrális momentum kiszámítása van hátra. Ennek képlete :

$$\mu_{x,y} = \frac{\sum_1^n (x_i - x_0)(y_i - y_0)}{n}$$

Az  $(x_i - x_0) \cdot (y_i - y_0)$  szorzatokat a 3. és 4. oszlopok összeszorzása révén kapjuk meg és a 7. oszlopba írjuk fel őket. Az összesorzásnál ügyelnünk kell az előjelre is, mely a szorzó és szorzandó előjelétől függően lehet pozitív és negatív is. A szorzatokat ezután előjelük szerint összegezzük és a végösszeget elosztjuk a tagok számával (20 db). A kapott szám nem más, mint a centrális momentum. Végül a centrális momentumot elosztjuk a két négyzetes középeltérés szorzatával, ami által megkapjuk magát a korrelációs tényezőt. A jelen példán a korrelációs tényező 0,9864-nek adódott ki, ami azt jelenti, hogy a két komponens között ebben a szakaszban a kapcsolat oly szoros, hogy majdnem függvénykapcsolatnak tekinthető. Összehasonlítás céljából kiszámítottuk az első és harmadik szakaszra vonatkozó korrelációs tényezőket is. A számítás fenti példánkkal analóg módon történt. A kis  $Al_2O_3$ -tartalmú szakaszban (30,5—39,5%) a korrelációs tényező 0,4363-nak, a nagy  $Al_2O_3$ -tartalmú szakaszban (58,5—64,5%) pedig 0,6244-nek adódott ki.

Láthatjuk tehát, hogy a két szakaszban az  $Al_2O_3$ - és  $SiO_2$ -tartalom kapcsolata sokkal lazább, mint a középső szakaszban. Ezekből a tényekből aztán fontos földtani és geokémiai következtetéseket lehet levonni.

Külön hangsúlyoznunk kell, hogy a korrelációs tényezőt a földtanban csak a kapcsolat jellegének megismerése — tehát a korrelációs diagram megszerkesztése után számíthatjuk ki. A földtani jelenségek bonyolult, komplex volta miatt a kapcsolatok jellege ugyanazon képződményen belül is megváltozhat, mint ahogy azt fenti példánkon is láttunk. E jellegváltozások statisztikus felismerése lehetőséget ad a földtani jelenségek teljesebb megismerésére. Ha viszont a korrelációs tényező számításakor a jellegváltozásokat nem vesszük figyelembe, úgy az egész fáradságos számtani művelet hiábavaló, a belőle levont következtetések pedig helytelenek lesznek.

Végezetül meg kell még említenünk, hogy a fenti példánkon tulajdonképpen „súlyponti-korrelációs tényezőt” számoltunk, mert nem egyes önálló számadatokkal, hanem több ezer adat átlagából kiadódó „súlyponti” adatokkal dolgoztunk (a fenti példán 1589 elemzési adat). Ez a módszer a számítások nagymérvű leegyszerűsítését tette lehetővé. Egyébként önálló pontpárok esetén is ugyanez a számítás módja, csupán a tagok számának rendkívüli megnövekedése (37 helyett 1589 db adat-pár) teszi ilyenkor a számítást igen hosszadalmassá.

A korrelációs számításnál nemcsak két tulajdonság kapcsolatát lehet meghatározni, hanem arra is lehetőség van, hogy egyszerre több (3—6) tulajdonság összefüggését is megvizsgáljuk.

Ez az ún. többszörös korreláció már jóval bonyolultabb számításokat igényel, mint az egyszerű kétváltozós korrelációs számítás, ezért most nem is foglalkozunk vele.

A matematikai statisztika módszereit tehát a felsorolt példák tanúsága szerint könnyen alkalmazhatjuk a földtanban. A számítások elvégzése nem igényel különösebb matematikai előképzettséget. Megfelelő számológépek segítségével a számításokat igen könnyen és gyorsan lehet elvégezni. Hazai vonatkozásban a statisztikai módszerek széleskörű alkalmazása nagymértékben elősegítheti földtani megismerésünk elmélyítését.



A SiO<sub>2</sub> tartalom eloszlási táblázata

1. táblázat

A SiO <sub>2</sub> tartalom osztályai	Elemzések száma	Gyakorisági %	1 × 3 szorzata	Kumulatív gyakorisági %	Eltérés a számtani közép-től $X_i - X_0$	Eltérés négyzete $(X_i - X_0)^2$	3 × 6 szorzata $W(X_i - X_0)^2$
1.	2.	3.	4.	5.	6.	7.	8.
0,0— 1,0	3	0,12	0,060	0,12	— 7,7	59,29	7,11
1,0— 2,0	168	6,69	10,035	6,81	— 6,7	44,89	300,31
2,0— 3,0	383	15,25	38,125	22,06	— 5,7	32,49	495,47
3,0— 4,0	313	12,47	43,645	34,53	— 4,7	22,09	275,46
4,0— 5,0	210	8,36	37,620	42,89	— 3,7	13,69	114,45
5,0— 6,0	175	6,97	38,335	49,86	— 2,7	7,29	50,81
6,0— 7,0	173	6,89	44,785	56,75	— 1,7	2,89	19,91
7,0— 8,0	137	5,45	40,875	62,20	— 0,7	0,49	2,67
8,0— 9,0	116	4,62	39,270	66,82	+ 0,3	0,09	0,42
9,0—10,0	109	4,34	41,230	71,16	+ 1,3	1,69	7,33
10,0—11,0	83	3,30	34,650	74,46	+ 2,3	5,29	17,46
11,0—12,0	81	3,23	37,145	77,69	+ 3,3	10,89	35,17
12,0—13,0	86	3,43	42,875	81,12	+ 4,3	18,49	63,42
13,0—14,0	76	3,03	40,905	84,15	+ 5,3	28,09	85,11
14,0—15,0	75	2,99	43,355	87,14	+ 6,3	39,69	118,67
15,0—16,0	49	1,95	30,225	89,09	+ 7,3	53,29	103,92
16,0—17,0	48	1,91	31,515	91,00	+ 8,3	68,89	131,58
17,0—18,0	30	1,20	21,000	92,20	+ 9,3	86,49	103,79
18,0—19,0	19	0,76	14,060	92,96	+10,3	106,09	80,63
19,0—20,0	18	0,72	14,040	93,68	+11,3	127,69	91,94
20,0—21,0	17	0,68	13,940	94,36	+12,3	151,29	102,88
21,0—22,0	5	0,20	4,300	94,56	+13,3	176,89	35,38
22,0—23,0	15	0,60	13,500	95,16	+14,3	204,49	122,69
23,0—24,0	12	0,48	11,280	95,64	+15,3	234,09	112,36
24,0—25,0	9	0,36	8,820	96,00	+16,3	265,69	95,65
25,0—26,0	12	0,48	12,240	96,48	+17,3	299,29	143,66
26,0—27,0	10	0,40	10,600	96,88	+18,3	334,89	133,96
27,0—28,0	8	0,32	8,800	97,20	+19,3	372,49	119,20
28,0—29,0	9	0,36	10,260	97,56	+20,3	412,09	148,35
29,0—30,0	9	0,36	10,620	97,92	+21,3	453,69	163,33
30,0—31,0	8	0,32	9,760	98,24	+22,3	497,29	159,13
31,0—32,0	3	0,12	3,780	98,36	+23,3	542,89	65,15
32,0—33,0	1	0,04	1,300	98,40	+24,3	590,49	23,62
33,0—34,0	2	0,08	2,680	98,48	+25,3	640,09	51,21
34,0—35,0	5	0,20	6,900	98,68	+26,3	691,69	138,34
35,0—36,0	4	0,16	5,680	98,84	+27,3	745,29	119,25
36,0—37,0	8	0,32	11,680	99,16	+28,3	800,89	256,28
37,0—38,0	5	0,20	7,500	99,36	+29,3	858,49	171,70
38,0—39,0	6	0,24	9,240	99,60	+30,3	918,09	220,34
39,0—40,0	4	0,16	6,320	99,76	+31,3	979,69	156,75
40,0—41,0	2	0,08	3,240	99,84	+32,3	1043,29	83,46
41,0—42,0	—	—	—	—	—	—	—
42,0—43,0	2	0,08	3,400	99,92	+34,3	1176,49	94,12
43,0—44,0	2	0,08	3,480	100,00	+35,3	1246,09	99,69
$\Sigma = 2510$		100 %	$\Sigma = 823,070$	100 %			$\Sigma = 4922,13$

Számítási közép = 8,2%

Diszperzió = 49,22

Medián = 6,0%

Négyzetes közép-eltérés = ± 7,02%

Módus = 2,5%

A bauxit  $Al_2O_3$ - és  $SiO_2$ -tartalmára vonatkozó korrelációs együttható számításai táblázata

III. táblázat

$Al_2O_3$ - tartalom nagyssága $X_i$	$SiO_2$ - tartalom nagyssága $Y_i$	$Al_2O_3$ számítási való eltérés $X_i - X_0$	$SiO_2$ középtől eltérés $Y_i - Y_0$	$Al_2O_3$ számítási való eltérés négyzete $(X_i - X_0)^2$	$SiO_2$ középtől eltérés négyzete $(Y_i - Y_0)^2$	$Al_2O_3$ és $SiO_2$ eltéréseinek szorzata $(X_i - X_0) \cdot (Y_i - Y_0)$
1.	2.	3.	4.	5.	6.	7.
39,5	33,8	-9,5	+18,2	90,25	331,24	-172,90
40,5	32,9	-8,5	+16,3	72,25	265,69	-138,55
41,5	29,5	-7,5	+13,9	56,25	193,21	-104,25
42,5	25,9	-6,5	+10,3	42,25	106,09	-66,95
43,5	25,7	-5,5	+10,1	30,25	102,01	-55,55
44,5	23,3	-4,5	+7,7	20,25	59,25	-34,65
45,5	20,1	-3,5	+4,5	12,25	20,25	-15,75
46,5	18,5	-2,5	+2,9	6,25	8,41	-7,25
47,5	15,4	-1,5	-0,2	2,25	0,04	+0,30
48,5	14,7	-0,5	-0,9	0,25	0,81	+0,45
49,5	12,3	+0,5	-3,3	0,25	10,89	-1,65
50,5	11,1	+1,5	-4,5	2,25	20,25	-6,75
51,5	10,8	+2,5	-4,8	6,25	23,04	-12,00
52,5	10,3	+3,5	-5,3	12,25	28,09	-18,55
53,5	7,4	+4,5	-8,2	10,25	67,24	-36,90
54,5	6,0	+5,5	-9,6	30,25	92,16	-52,80
55,5	5,3	+6,5	-10,3	42,25	106,09	-66,95
56,5	3,7	+7,5	-11,9	56,25	141,61	-89,25
57,5	3,1	+8,5	-12,5	72,25	156,25	-106,25
58,5	2,7	+9,5	-12,9	90,25	166,41	-122,55
980,0	312,5			665,00	1899,06	-1108,75

Tagok száma  $n = 20$  $Al_2O_3$  számítási közepe = 49,0% $SiO_2$  számítási közepe = 15,6%

$$\sigma_x = \sqrt{\frac{\sum_1^n (x_i - x_0)^2}{n}} = \sqrt{\frac{665,0}{20}} = \sqrt{33,25} = 5,77$$

$$\sigma_y = \sqrt{\frac{\sum_1^n (y_i - y_0)^2}{n}} = \sqrt{\frac{1899,06}{20}} = \sqrt{94,95} = 9,74$$

$$\mu_{xy} = \frac{\sum_1^n (x_i - x_0) \cdot (y_i - y_0)}{n} = -\frac{1108,75}{20} = -55,437$$

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \cdot \sigma_y} = -\frac{55,437}{5,77 \cdot 9,74} = -\frac{55,437}{56,1998} = -0,9864$$

## IRODALOM — ЛИТЕРАТУРА — LITERATURE

- I. Bogárdi János: Korreláció számítás és alkalmazása a hidrológiában. Akadémiai Kiadó, 1952. — 2. Jordán K.: Matematikai statisztika. Bpest, 1927. Athenaeum kiadás. — 3. Jordán K.: A korreláció számítása. Magyar Statisztikai Szemle kiadványai, 1941. Bpest. — 4. Krumbain, W. C. and Pettijohn, F. J.: Manual of Sedimentary Petrography. New York, 1938. — 5. Реньи А.: Valószínűség számítás. Tankönyvkiadó, 1954. — 6. Рыжов, П. А. Геометрия недр, Москва 1952. — 7. Сзентмáртону Т.: Matematikai statisztika a mészaki gyakorlatban. Bpest, 1950. A mérnöki továbbképző intézet. — 8. Trask, P. D.: Origin and Environment of Source Sediments of Petroleum. Houston Texas, 1932.

**Применение статистических способов в геологии**

Г. БАРДОШИ

Автор излагает способы математической статистики, пригодные для использования в геологии. Излагаются исходные понятия статистики на примерах, взятых из области геологии. Так он занимается таблицами распределения статистической совокупности, частотами и диаграммами распределения (статистический полигон, гистограмма, кумулятивная кривая и т. п.). В дальнейшем автор излагает важнейшие параметры, определяющие дисперсию и асимметрию статистической совокупности. Наконец излагается способ корреляционного анализа для двух вариантов. По всем способам автор приложил конкретные счетные примеры из практики геологии.

**Application of Statistical Methods in Geology**

by GY. BÁRDOSSY

The author discusses some methods of mathematical statistics and their application in geology. At first he deals with the frequency distribution of a statistical series. He shows several methods of graphic presentation of statistic data, as frequency polygons, histograms, cumulative curves and their mutual relations. Afterwards he discusses the most important parameters like the measures of central tendency, measures of dispersion and measures of skewness. On the end he gives a short account of correlational analysis with two variables. For all the methods he gives computational examples from the territory of geological practice.