

FÉNYLYUKKÁRTYÁS ADATNYILVÁNTARTÓ RENDSZER OLAJKUTAK ADATAINAK NYILVÁNTARTÁSÁRA ÉS STATISZTIKAI SZÁMÍTÁSOK VÉGZÉSÉRE

Kassay Árpád

Kőolaj- és Földgázbányászati Ipari Kutató Laboratórium
Nagykanizsai Üzemegysége

A Magyarországon működő olajkutak jelentős része segédgázos működtetésű. Ezek üzemének javításához, módosításához olyan adatok keltenek, amelyek a már meglévő kutak üzemeltetésére vonatkoznak, ugyanakkor szükségesek a kutak egyéb adatai is, így pl. a kutak mélysége, olaj-fajsúly, olaj-viztartalom, nyomások stb. Ezen adatoknak a kb. 800 kutra történő feldolgozása /kutanként kb. 80 adattal/ tekintélyes mennyiségű munkát vesz igénybe manuális feldolgozás esetén. A megoldást tehát egy olyan módszer jelentheti, amely egyaránt alkalmas a nyilvántartásra, a gyors visszakeresésre, és az adatok egyszerű statisztikai feldolgozására is.

Ilyen célra általában a számítógépes adatnyilvántartást szokták felhasználni, amely módszernek vannak előnyei és hátrányai is.

Előnye, hogy rendkívül gyors módszer, az alapadatokból megfelelő programok segítségével leszámaztatott adatok kaphatók, a statisztikai feldolgozás igen egyszerűen megoldható stb.

Hátránya többek között az, hogy a jelen helyzetben a számítógépek nehezen hozzáférhetők, általában távol vannak a felhasználás helyétől, ebben az esetben pedig telex vagy terminál-kapcsolatot igényelnek, végül pedig a nagy teljesítményű számítógépek igénybevétele ilyen - viszonylag kis számú - adathalmaz esetén eléggé költséges.

Fenti megfontolások alapján úgy döntöttünk, hogy bár a számítógépes adatfeldolgozás távlati lehetőségeit fenntartjuk, jelenleg azonban egyszerűbb, jól használható módszerre van szükségünk; így a már jól bevált fénylyukkártyás adatnyilvántartáshoz folyamodtunk, amelynek előnyei közismertek.

Figyelembe kell venni, hogy a fénylyukkártyás adattárolási rendszer tulajdonképpen nem nyilvántartási, még kevésbé statisztikai adatfeldolgozási műveletekre szolgál, hanem fő célját tekintve információkereső eszköz. Ennek megfelelően az egyes fénylyukkártyák által képviselt szempontok /ismérvek/ csupán a mindenkori, tárolt adatok jellemzésére, megjelölésére szolgálnak. Következésképpen a fénylyukkártyák segítségével elvileg csupán azt lehetne megállapítani, hogy

valamely meghatározott jellemzőkkel rendelkező adat létezik-e, és melyik az, ill. mekkora az e meghatározott jellemzőkkel rendelkező adatok száma.

Az egyes fénylyukkártyáknak megfeleltetett ismérvek /jellemzők/ lehetnek mennyiségi és minőségi jellegűek. Ennek megfelelően az egy-egy fénylyukkártyán lyukasztott pozíciók mennyisége annak a halmaznak számosságát jelzi, amely halmaz az adott mennyiségi vagy minőségi jellemzőkkel rendelkezik. Első közelítésre is világos tehát, hogy el kell különíteni egymástól a mennyiségi ismérvek, ill. a tárolt adatok mennyiségének fogalmát.

A minőségi ismérvekkel viszonylag könnyű dolgunk van, mivel itt nem merülhet fel más igény, mint a megfelelő minőségi ismérvvel /vagy ismérvek kombinációjával/ rendelkező egyedi adatok /pl. meghatározott kutak/ keresése, ill. az adott minőségi jellemzőknek megfelelő adatok /pl. kutak/ számának meghatározása. Az egy-egy minőségi csoportba tartozó minőségi jellemzők lehetnek igen egyszerű, "igen-nem" szerkezetűek, mint pl. esetünkben az olyan kérdés, hogy egy adott kut működésének jellege időszakos-e vagy nem, és amennyiben időszakos, akkor van-e beépítve béléscső automata vagy sem. Ilyen esetben nyilvánvalóan két-két eset lehetséges, és ezek rögzítésére általában egy-egy fénylyukkártya elegendő. Egy minőségi ismérv-csoportba azonban nagyobb számú /egymást általában kölcsönösen kizáró/ olyan minőségi ismérvek is tartozhatnak, amelyek egyenként mind "igen-nem"-mel megválaszolhatók /pl. a kut kőzetadottságai/.

Más a helyzet a mennyiségi /számszerű/ ismérveknél. Léteznek természetesen olyan esetek, amikor a mennyiségi ismérv egyben minőségi, mivel nem a számszerű /digitális/ adatok folytonos soráról van szó, hanem néhány, minőségileg is meghatározó számszerű adatról. Így pl. alkalmazott villamos erőgépek esetében elképzelhető vagy lehetséges, hogy ezen erőgépek teljesítménye csupán ötféle és az adott, számszerű teljesítmény-mutató egyedileg egyetlen osztályt jelent. Más esetekben viszont a mennyiségi adatok lényegében folyamatos sort alkotnak, és nyilvánvalóan értelmetlen vagy semmitmondó lenne /önkényesen meghatározott/ mennyiségi egységenként /pl. méterenként/ egy-egy fénylyukkártyának megfeleltetni e mennyiségi ismérveket. Mennyiségi ismérv pl. a kutak mélysége. Nyilvánvalóan értelmetlen lenne /amellett, hogy ez is csupán önkényes megoldás, hiszen nagyobb vagy kisebb egységet is választhattunk volna/ méterenként meghatározni a kutmélység mennyiségi ismérvét.

A mennyiségi ismérvek esetében ezért általában arra kényszerülünk /és éppen ezt a kényszerből adódó lehetőséget használjuk ki a későbbiekben/, hogy a mennyiségi ismérvekkel jellemezhető adatokat osztályokba soroljuk, tehát, hogy a mennyiségi ismérveket nem egyedileg, hanem alsó és felső határok között határozzuk meg. Így pl. a kutak mélységénél az előforduló legkisebb mélységtől az előforduló vagy elképzelhető legnagyobb mélységig terjedő, számszerű adatokkal kifejezhető tartományt alsó és felső határokkal meghatározott résztartományokra, intervallumokra oszthatjuk, amelyek azután mint osztályok jelentik a fénylyukkártyának megfeleltetett mennyiségi ismérvet.

Az osztályok mennyiségi tartománya lehet azonos és változó. Változó terjedelmű osztályokat /intervallumokat/ alkalmaz a statisztika olyan esetekben, amikor a mennyiségi változás egyenlőtlenül fejez ki minőségi változásokat, és ezért célszerű, ha a határokat e minőségi változásoknak megfelelően szabják meg /pl. üzem nagyság-statisztikák/. Más esetekben azonban semmi sem szól az egyenlő intervallumok ellen, sőt - amint látni fogjuk - éppen a megalapozott terjedelmű, egyenlő intervallumokra osztott mennyiségi adatsorok adják meg a lehetőséget arra, hogy a fénylukkártyákon tárolt adatokkal műveleteket végezhessünk, ami változó intervallumu osztályok esetében nem lenne lehetséges.

Az olajkutak itt említett nyilvántartásánál gyakran van szükség ilyen statisztikai feldolgozási műveletekre. Nem csupán azt kell megállapítani például, hogy mely kutak tartoznak mélységük szempontjából egy adott osztályba /intervallumba/, ill. hány darab kut tartozik az egyes osztályokba /intervallumokba/, hanem pl. azt is, hogy mennyi az összes kutak átlagos mélysége és mekkora az átlagos mélységtől való négyzetes eltérés. Az itt ismertetett eljárás e műveletek fénylukkártyás nyilvántartás alapján való elvégzésének lehetőségét mutatja be.

Az osztályok optimális terjedelmének meghatározására a szakirodalom* az alábbi közelítő képlet alkalmazását javasolja:

$$h = 0,08 \sqrt{x_{\max} - x_{\min}}$$

ahol

h - az osztályterjedelem

x_{\max} és x_{\min} - az osztályba sorolandó adatok maximális és minimális értéke.

Mivel az egyenlet közelítő, az egyes osztályok terjedelmét a számítás után úgy kerekíthetjük, hogy az osztályok száma 10 és 20 között legyen.

Azt az értéket, amelyik az osztály középértékének felel meg és amely az alsó és felső határral meghatározott osztályba tartozó összes értéket képviseli, osztályindex-nek nevezzük. Az osztályindexet a számtani közép ismert képlete alapján számítjuk ki:

$$x_i = \frac{x_{i,h} + x_{i,d}}{2}$$

ahol $x_{i,h}$ és $x_{i,d}$ az a legnagyobb ill. legkisebb érték, amelyet az i -edik osztályba be lehet sorolni.

* FELIX, M.; BLAHA, K.: Matematikai statisztika a vegyiparban. Bp. Műszaki Könyvkiadó, 1964.

Az osztálybasorolás az alábbiak szerint történik: legyen a legkisebb és a legnagyobb kutmélység értéke 566,5 m és 1985 m. Ekkor

$$h = 0,08 / 1985,0 - 566,5 / = 113,48$$

Az eredményt 100-ra kerekítjük és az egyes osztályokat elhatároljuk:

Osztályhatárok	Oszt.index	Osztályhatárok	Oszt.index
550 - 650	600	1350 - 1450	1400
650 - 750	700	1450 - 1550	1500
750 - 850	800	1550 - 1650	1600
850 - 950	900	1650 - 1750	1700
950 - 1050	1000	1750 - 1850	1800
1050 - 1150	1100	1850 - 1950	1900
1150 - 1250	1200	1950 - 2050	2000
1250 - 1350	1300		

Látható, hogy az osztálybasorolás alapján az osztályok száma valóban 10 és 20 közötti, esetünkben 15.

Ezek után már nincs más tennivalónk, csak az egyes osztályindexeknek és/vagy az alsó és felső osztályhatároknak megfelelő értékekkel feliratozzuk a kártyákat, és a megfelelő osztálybasorolás alapján az osztályba tartozó adatoknak megfelelő kutsorszámot belyukasztjuk a kártyába.

Az ilyen típusú feldolgozás nagy előnye az, hogy igen egyszerű módon teszi lehetővé a statisztikai feldolgozást a fénylyukkártyák segítségével úgy, hogy az eredeti, pontos adatokra, amelyek az egyes kutak adatainál találhatók, nincs is szükségünk. Itt csak az osztályindexek alapján kapott gyakoriságot, vagyis az egyes fénylyukkártyákon belyukasztott lyukak számát vesszük figyelembe az azonos adatscsoporton belül. Ezt az egyes osztályok abszolút gyakoriságának nevezzük és n_i -vel jelöljük. A megfelelő abszolút gyakoriságok értékeiből kaphatjuk meg a relatív gyakoriságot, ha azt alosztjuk az összes osztályba sorolt, megfelelő adatok számával.

A gyakorlatban gyakran előfordul, hogy bizonyos területen található kutak paramétereinek átlagára van szükség, de - a bizonytalansági tényezők számításba vétele érdekében - az átlagtól való eltéréssel /szórásnégyzet/ - együtt.

Az n_i abszolút gyakoriság segítségével a számtani középértéket és a szórásnégyzetet a következő képletekkel tudjuk kiszámítani:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m x_i n_i$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^m /x_i - \bar{x}^2 n_i$$

ahol n_i - az i -edik osztály gyakorisága, ha $i = 1, 2, \dots$

x_i - az i -edik osztály osztályindexe

n - az adott ismerv összetartozó adatainak száma

m - az osztályok száma.

A fenti képletek alapján, amennyiben rendelkezésünkre áll egy programozható számológép, akkor igen egyszerű a jellemzők kiszámítása, amelyet tovább nem is kell részleteznünk. Vállalatunknál ezt a feladatot egy Hewlett-Packard 9810A típusú programozható asztali számológép segítségével végeztük el.

Azok számára, akik nem rendelkeznek ilyen géppel és számolni sem nagyon szeretnek, ismertetünk egy egyszerű módszert, amely csak egész számokkal dolgozik és így igen egyszerűen, szinte fejben kiszámolható az eredmény.

A módszer matematikai alapja a következő: vezessünk be egy új változót, v_i -t az x_i osztályindex helyett, amely utóbbi gyakran több számjegyet tartalmazó tizedes tört. A v_i meghatározása - osztályonként - az alábbi képlettel történik:

$$v_i = \frac{x_i - x_0}{h}$$

ahol x_0 - egy megfelelően megválasztott állandó

h - az osztály terjedelme.

Gyakorlati okokból célszerű, ha x_0 -nak a legkisebb és legnagyobb osztályindexek közéjéhez közeli osztályindexek valamelyikét vesszük.

Az új változó értékei mindig egész számok. A számított jellemzőkből az eredeti változó átlagának és szórásának értéke a következőképpen számítható:

$$\bar{x} = h \bar{v} + x_0$$

$$\sigma_x = h \sigma_v$$

$$\text{ahol } \bar{v} = \frac{1}{n} \sum_{i=1}^m v_i n_i$$

$$\text{és } \sigma_v^2 = \frac{1}{n} \sum_{i=1}^m v_i^2 n_i - \bar{v}^2$$

Segítségképpen az alábbi példát közöljük:

/Az új változó értéke $x_0 = 1400$; $h = 100$./

Oszt. index x_i	Gyakoriság n_i	Új változó v_i	Számítás	
			$v_i n_i$	$v_i^2 n_i$
600	1	-8	-8	64
700	2	-7	-14	98
800	4	-6	-24	144
900	5	-5	-25	125
1000	7	-4	-28	112
1100	6	-3	-18	54
1200	8	-2	-16	32
1300	10	-1	-10	10
1400	14	0	0	0
1500	11	1	11	11
1600	9	2	18	36
1700	7	3	21	63
1800	6	4	24	96
1900	2	5	10	50
2000	1	6	6	36
Összesen	93		-53	921

Az $x_1 = 600$, tehát

$$v_1 = \frac{600 - 1400}{100} = -8$$

$x_2 = 700$, így $v_2 = \frac{700 - 1400}{100} = -7$, és így tovább.

Gyakorlatilag ezt úgy végezzük, hogy a táblázatban az x_0 kiválasztott értékének sorába mindig 0-t írunk. Az ennél kisebb értékek esetén a táblázatnak megfelelően mindig osztályonként eggyel kisebb negatív számot, a nagyobbaknak eggyel nagyobb pozitív számot adunk az x_0 -tól kiindulva. A szorzatok kiszámítása fejből történhet. Ezután elvégezzük az összegezést, majd kiszámítjuk a jellemzőket az alábbi példának megfelelően:

$$\bar{v} = \frac{1}{n} \sum_{i=1}^m v_i n_i = \frac{1}{93} /-53/ = -0,5699$$

majd az alábbi részszerítést végezzük el:

$$\frac{1}{n} \sum_{i=1}^m v_i^2 n_i = \frac{1}{93} /921/ = 9,9032$$

Ennek segítségével:

$$\sigma_v^2 = \frac{1}{n} \sum_{i=1}^m v_i^2 n_i - \bar{v}^2 = 9,9032 - 0,3248 = 9,5784$$

Az eredeti változók szerint tehát a következőket kapjuk:

$$\bar{X} = h \bar{v} + x_0 = 100 /-0,5699/ + 1400 = \underline{1343 \text{ m}}$$

$$\sigma_x^2 = h^2 \sigma_v^2 = 100^2 \cdot 9,5784 = 95.784$$

$$\sigma_x = \sqrt{95.784} = 309,49 \text{ m}$$

Ezek az adatok minimális eltéréssel megegyeznek az összes adat feldolgozásával végzett hosszadalmas feldolgozás adataival.

A fentiek alapján látható, hogy a fénylyukkártyás adatnyilvántartás nem csupán információátvitelre és -keresésre, hanem statisztikai számítások végzésére is alkalmas. A módszer igen egyszerű és igen pontos eredményeket szolgáltat. Ezzel a módszerrel a fénylyukkártyás adatnyilvántartás hatékonyságát nagymértékben meg lehet növelni.

A módszer természetesen olajkutak nyilvántartása esetében sem csak segédgáz kutak nyilvántartására és adataik számítására alkalmas, hanem minden olyan esetben, amikor hasonló mennyiségi jellemzőkkel rendelkező sokaságok statisztikai számítására van szükség.

KASSAY Árpád: Fénylyukkártyás adatnyilvántartó rendszer olajkutak adatainak nyilvántartására és statisztikai számítások végzésére

Mint ismeretes, a fénylyukkártyák dokumentum- és adatgyűjtemények különböző ismérvek /szempontok/ szerinti nyilvántartását és visszakeresését teszik lehetővé. Ezek az ismérvek minőségi és mennyiségi vonatkozásúak lehetnek. Az utóbbiak gyakran meghatározott kezetek közé eső mennyiségeket reprezentálnak. Ebben az esetben, a fénylyukkártyákon nyilvántartott adatok alapján, nemcsak az egységek keresésére van lehetőség, hanem az említett mennyiségi ismérvek átlagértékeinek és szóródásának elemzésére is.

A cikk a hazai olajkutak példáján mutatja be az adatkeresés és statisztikai elemzés, számítógép nélküli, gyorsított módszerét és ezzel a fénylyukkártyák ujszerű felhasználásának lehetőségét.

A számjegyes adatokat matematikai módszerekkel optimális számú osztályokba sorolták, és az egyes osztályokba tartozó adatok középértékének kiszámításával ún. osztályindexeket alkottak. A kártyákat az osztályindexeknek megfelelő értékekkel feliratozták, és az azonos osztályba tartozó adatoknak /osztályindexeknek/ megfelelő kut sorszámát lyukasztják a megfelelő kártyákra.

A minőségi adatok nyilvántartására az "igen-nem" keresési elvet alkalmazták, vagyis az egymást kölcsönösen kizáró adatok esetében, csak az egyik adatnak /ismérvnek/ nyitnak kártyát.

00

KASSAY, Á.: Registration and measuring of data of oil-wells on punched cards for visual selection

Optical coincidence /peek-a-boo/ cards - as it is well known - enable multi-aspect registration and retrieval of document- and data-collection. These aspects could be qualitative and quantitative as well. The latter ones are often representing quantities between given limits. In this case the data registered on punched cards enable not only to search certain entries but also to analyze the average value and scattering of the quantitative aspects mentioned.

By the example of domestic oil-wells the article illustrates a modern application possibility of peek-a-boo cards for high speed data search and statistical analysis without computers.

Numerical data were classified by mathematical methods into optimal numbers of classes and class indexes were made by calculation of the mean of data pertaining to single classes. The cards were labelled according to the values of class indexes then the numbers of oil-wells - respective to data /class indexes/ included into common classes - were punched.

For registering qualitative data the "yes-no" search method was applied i.e. in case of disqualifying data only one aspect /characteristics/ was put on card.

.□.

КАШШАИ, А.: Хранение данных нефтяных скважин и проведение статистических расчетов с помощью системы хранения данных на просвечивающихся перфокартах

Как известно, просвечивающиеся перфокарты позволяют хранение документов и данных, их поиск по различным свойствам (показателям). Эти свойства могут быть как качественными, так и количественными. Последние часто представляют собой величины, находящиеся в определенном интервале. В этом случае существует возможность не только для поиска отдельных единиц, но и для анализа средней величины количественных показателей и их дисперсии.

На примере венгерской сети нефтяных скважин в статье описывается ускоренный метод поиска данных и статистического анализа без использования ЭВМ, и тем самым представляется новая возможность использования просвечивающихся перфокарт.

По количественным показателям скважины — с помощью математических методов — были разбиты на классы, число содержащихся которых оптимальное. Вычислением средних величин данных, находящихся в отдельных классах, были определены т.ч. индексы классов. На перфокартах сверху пишутся величины, соответствующие индексу данного класса, и на соответствующих картах пробиваются номера скважин, соответствующих данным одного и того же класса (одному и тому же индексу класса).

При хранении качественных данных использовался метод поиска "да-нет", т.е. в случае взаимно исключающих друг друга данных перфокарта подводится только для одного из данных (свойств).

KASSAY, Á.: Datenverarbeitungssystem mit Sichtlochkarten für Registrierung der Daten von Erdölquellen und Verrichtung von statistischen Berechnungen

Wie bekannt, ermöglichen die Lochkarten die aufgrund verschiedener Kennzeichen /Gesichtspunkte/ erfolgende Registrierung und Recherche von Dokumenten- und Datensammlungen. Die Kennzeichen können von quantitativer und qualitativer Art sein. Die quantitativen Kennzeichen repräsentieren häufig innerhalb bestimmte Grenzen fallende Quantitäten. In diesem Fall ist aufgrund der auf den Sichtlochkarten registrierten Daten nicht nur die Recherche der Einheiten, sondern auch die Analyse der Mittelwerte der erwähnten quantitativen Kennzeichen und ihrer Streuung möglich.

In der Arbeit wird am Beispiel der ungarischen Erdölquellen die beschleunigte Methode ohne Rechner der Datenrecherche und der statistischen Analyse dargestellt und damit auf eine neue Einsatzmöglichkeit der Sichtlochkarten hingewiesen.

Die ziffernmässigen Daten wurden mit mathematischen Methoden in eine optimale Zahl von Klassen eingereiht und durch die Ermittlung des Mittelwertes der in die einzelnen Klassen gehörenden Daten wurden sog. Klassenindizes gebildet. Die Karten wurden mit den den Klassenindizes entsprechenden Werten beschriftet und die Ordnungszahl der in die identische Klasse gehörenden Daten /Klassenindizes/ entsprechenden Erdölbrunnen wurden auf die entsprechenden Karten ge-
locht.

Für die Evidenz der qualitativen Daten wurde das "ja-nein"-Rechercheprinzip angewendet, das heisst im Fall von einander gegenseitig ausschliessenden Daten wird nur für eine der Angaben /Kennzeichen/ eine Karte fertiggestellt.

