Весна Јевремовић

О НЕКИМ МАЊЕ УОБИЧАЈЕНИМ НАЧИНИМА ГРАФИЧКОГ ПРЕДСТАВЉАЊА СТАТИСТИЧКИХ ПОДАТАКА

Уобичајени део статистичке обраде података је њихово графичко представљање. Оно помаже при интерпретацији резултата јер визуелно добијена информација олакшава памћење или поређење. Правилно презентирани и интерпретирани статистички подаци нам помажу и у многим свакодневним одлукама. Статистика тако постаје део опште културе. Статистичари би наравно волели да тај процес буде бржи и обухвати још већи део становништва. Као и код свих других наука и у статистици се паралелно са напретком, јавља и нестајање са сцене неких појмова и метода које су раније биле коришћене, пре но што су рачунари преузели „главну улогу" у обради података. Подсетићемо овде на неколико таквих појмова. Наш избор је везан за оне у чијем доказивању или интерпретацији користимо одређена знања из геометрије.

Најпре наводимо две могућности табличног представљања података. Нека имамо укупно n података.

(а) Вредности посматране појаве (обележја) бележене су појединачно. Нека су то бројеви (као што је уобичајено, претпостављамо да су вредности обележја дате у растућем поретку): $x_1 < x_2 < \cdots < x_p$:

вредности обележја	x_1	x_2	$\cdots$	x_p
фреквенције	n_1	n_2	$\cdots$	n_p

$$n_1 + n_2 + \cdots + n_p = n$$

Табела 1.

(б) Вредности посматране појаве (обележја) бележене су по интервалима (чест случај уколико је укупан број података велики):

вредности обележја	$(-\infty, a_1)$	$[a_1, a_2)$	$\cdots$	$[a_{p-1}, \infty)$
фреквенције	n_1	n_2	$\cdots$	n_p

$$n_1 + n_2 + \cdots + n_p = n$$

Табела 2.

Могуће је притом да први или последњи интервал такође буду коначни, као и да се интервали узимају затворени са десне, а отворени са леве стране. Груписањем података у Табели 1, добијамо Табелу 2, а ако у Табели 2 узмемо из сваког интервала по један број као представника тог интервала (нпр. средиште интервала) онда добијамо Табелу 1.

Поларни дијаграм

Код представљања података који садрже цикличну компоненту (али не само њих) као што су подаци сређени по годишњим добима, по месецима, часовима, . . . , можемо применити поларни дијаграм. Податке представљамо на полуправим које све полазе из координатног почетка (обично су углови између суседних полуправих једнаки) и тако добијамо представу о томе како се посматрано обележје развија у времену. Линеарна интерполација се лако остварује на дијаграму.

Пример. Подаци о броју незапослених (у хиљадама) за неку област дати су за период од две године:

месец	јан	феб	мар	апр	мај	јун	јул	авг
број незапослених	1121	1100	1075	1050	1041	1040	1095	1160

сеп	окт	нов	дец	јан	феб	мар	апр
1280	1340	1330	1328	1356	1342	1313	1290

мај	јун	јул	авг	сеп	окт	нов	дец
1260	1236	1257	1306	1424	1490	1473	1470

Поларни дијаграм у овом примеру конструишемо тако што из координатног почетка конструишемо 12 полуправих, за сваки месец по једну, тако да између суседних буде угао од 30°. Спајајући одговарајуће тачке на овим полуправим добијамо једну спиралу која нам показује како се мења број незапослених у посматраном периоду, и омогућава да поредимо податке за исте месеце у различитим годинама. Испрекиданом линијом (в. слику) добијамо линеарну интерполацију и тако процењујемо број незапослених, на пример, средином једног месеца.

Сл. 1

Троугаони дијаграм

Ако се нека појава (обележје) може јавити у три облика (варијетета, модалитета), податке можемо представити троугаоним дијаграмом, користећи познату особину једнакостраничног троугла да је збир растојања произвољне тачке у унутрашњости троугла од његових страница константан.

Пример. Дати су подаци (у процентима) о производњи електричне енергије из хидроцентрала, термоцентрала и нуклеарних централа, за период од 20 година:

година	хидро	термо	нуклеарне
1961	40.00	39.00	21.00
1971	35.00	45.00	20.00
1976	30.00	45.00	25.00
1981	30.00	40.00	30.00

Троугаони дијаграм је дат на слици 2. Можемо пратити кретање тачака које одговарају појединим годинама у унутрашњости троугла. Тачка M_1 одговара 1961. години. Њена растојања од страница троугла су пропорционална фреквенцијама 40.00, 39.00, 21.00, односно једнаке $0.40h$, $0.39h$ и $0.21h$, где је h висина троугла.

Сл. 2

Лоренцова крива и параметар концентрације

Уобичајене нумеричке карактеристике обележја су узорачка средина и дисперзија. Од значаја је посматрати и неке друге карактеристике, зато што, на пример, ни узорачка средина ни дисперзија не дају одговор на питање да ли је укупна маса новца за личне дохотке у неком предузећу равномерно (релативно мало веома малих и веома великих плата) или неравномерно (већи број веома малих и веома великих плата) распоређена. Одговор на такво питање даје Лоренцова крива. Наравно да се она може применити и за неке друге податке: број изостанака ученика, број саобраћајних прекршаја, потраживања неког предузећа, ... При томе вредности обележја треба да буду позитивне, што се транслацијом може увек постићи. Нека су дати подаци (као што је уобичајено, претпостављамо да су вредности обележја дате у растућем поретку $0 < x_1 < \cdots < x_p$):

вредности обележја	x_1	x_2	$\cdots$	x_p
фреквенције	n_1	n_2	$\cdots$	n_p

$$n_1 + n_2 + \cdots + n_p = n$$

Лоренцова крива је полигонална линија која полази од координатног почетка $O(0,0)$ и спаја редом тачке $T_i(u_i, v_i)$, чије координате одређујемо по формулама

$$u_i = \frac{\sum\limits_{j=1}^{i} n_j}{n}, \quad v_i = \frac{\sum\limits_{j=1}^{i} n_j x_j}{\sum\limits_{j=1}^{p} n_j x_j}, \qquad i = 1, 2, \ldots, p.$$

Лоренцова крива се завршава у тачки $A(1,1)$. Ако је крива ближа дијагонали OA, расподела је равномернија, а што је крива удаљенија од дијагонале OA, расподела је више концентрисана на мање или веће вредности обележја.

Може се доказати да је Лоренцова крива конвексна или, што је у овом случају исто, да је коефицијент правца сваког одсечка Лоренцове криве већи од коефицијента правца претходног одсечка.

Пример. Конструисати Лоренцову криву за податке дате у табели:

вредности	80	120	250	440
фреквенције	140	50	110	200

Овде су тачке: $T_1(0.28, 0.084)$, $T_2(0.38, 0.130)$, $T_3(0.6, 0.337)$ и $T_4 = A(1,1)$. Лоренцова крива је дата на слици 3.

Сл. 3

Сл. 4

Индекс концентрације једнак је двострукој површини области ограничене дијагоналом OA и Лоренцовом кривом. Рачунамо га тако што израчунамо површину између Лоренцове криве и осе Ox, добијену вредност помножимо са 2 и одузмемо од 1. Што је индекс концентрације ближи јединици, концентрација је већа. За брзу процену индекса концентрације можемо узети двоструку дужину одсечка MN (слика 3). (У датом примеру је индекс концентрације једнак 0.32, док је дуж MN једнака 0.16). До овог резултата долазимо ако Лоренцову криву апроксимирамо луком одговарајуће кружнице. Површина кружног одсечка је приближно једнака дужи MB, види слику 4. Разлика је реда величине $0.04\,\alpha^3$, где је α половина централног угла за посматрани исечак. Докажите!

Одређивање мода расподеле

Мод је вредност обележја са фреквенцијом већом од суседних. Мод може, али не мора постојати. Ако постоји, не мора бити јединствен. Мод се лако одређује када су подаци дати појединачно, као у Табели 1.

Нека су подаци дати сређени по интервалима (класама) једнаке дужине и нека је конструисан хистограм на основу Табеле 3:

вредности обележја	$[a_0, a_1)$	$[a_1, a_2)$	$\cdots$	$[a_{p-1}, a_p)$
фреквенције	n_1	n_2	$\cdots$	n_p

$$n_1 + n_2 + \cdots + n_p = n$$

Табела 3.

Сл. 5

(Висине правоугаоника једнаке су одговарајућим фреквенцијама.)

Претпоставимо да постоји само једна модална класа (којој одговара правоугаоник већи од оба суседна). Мод M се тада рачуна по формули

$$M = L + \frac{\delta c}{\delta + \Delta},$$

где је $[L, D)$ модална класа (интервал), c дужина модалне класе, δ разлика фреквенције модалне класе и класе која је лево од ње, Δ разлика фреквенција модалне класе и класе која је десно од ње.

Сл. 6

Показаћемо како се мод може конструисати на хистограму. На слици 6, из сличности троуглова PQR и PST налазимо

$$\frac{EP}{RQ} = \frac{PF}{ST},$$

односно

$$\frac{M-L}{\delta} = \frac{D-M}{\Delta},$$

а како је $D = L + c$, после сређивања добијамо наведени резултат.

Наведени резултат има и следећу геометријску интерпретацију: апсциса максимума параболе која садржи средишта одсечака AQ, RS и TB једнака је M. Докажите!

ЛИТЕРАТУРА

[1] M. R. Spiegel, *Statistics*, McGraw-Hill Book Co, New York 1972.

[2] Ch. Leboeuf, J.-L. Roque, J. Geugand (eds.), *Cours de probabilites et de statistiques*, Paris 1981.

ОБАВЕШТЕЊЕ

УПРАВНИ ОДБОР И СКУПШТИНА ДРУШТВА МАТЕМАТИЧАРА СРБИЈЕ

У дане одржавања Републичког семинара '97 о настави математике и рачунарства одржане су и седнице органа Друштва.

Седница Управног одбора Друштва математичара Србије одржана је 10. јануара 1997. године. Управни одбор је размотрио и допунио припремљени извештај о раду Друштва у претходне две године.

Редовна скупштина Друштва математичара Србије одржана је такође 10. јануара 1997. године. Скупштина је усвојила извештај о раду Друштва у протеклом двогодишњем периоду. Извештај су поднели председник Друштва др Павле Младеновић и председник Управног одбора Друштва др Синиша Црвенковић. На Скупштини је за председника Друштва у наредном двогодишњем периоду поново изабран др Павле Младеновић.