分類　～ワケル博士～

愛知県立瑞陵高等学校　教諭　髙原　文規

問題

ワケル博士は都道府県の分類で頭を抱えています。

下の表は日本の都道府県のデータです。

X は㎢を単位とした面積、Y は千人を単位とした人口を表しています。

面積が平均の 7768.5 ㎢ 以上なのは 11 道県、人口が平均の 2724.6 千人以上なのは 12 都道府県で平均以上、平均未満を評価として分類すれば都道府県を

(面積平均以上, 人口平均以上), (面積平均以上, 人口平均未満),
(面積平均未満, 人口平均以上), (面積平均未満, 人口平均未満)

の 4 つの類型に分類することができます。

ワケル博士は、もう一つ

$A \times$ (面積) $+ B \times$ (人口) が C 以上, 未満

という評価を考えれば、上の 4 つの類型がそれぞれ 2 つに別れて 8 つの類型分類できると思ったが、いろいろ A, B, C を考えても、8 つのうち 1 つの類型に都道府県が 1 つも入らないので頭を抱えています。

ワケル博士に代わって、うまく 8 つの類型に都道府県が別れる A, B, C を考えるか、なぜうまくいかないのか説明してください。

ワケル博士は、上の方法ではうまくできなかったので、データをもう 1 種類増やして、第 1 次産業の生産額 Z(単位は 10 億円) を一緒に考えました。生産額が平均以上, 未満を評価に加えれば、8 つの類型に分類することは可能でした。
ところが、

$A \times$ (面積) $+ B \times$ (人口) $+ C \times$ (生産額) が D 以上, 未満

という評価を考えても、またしても 16 個の類型になりません。

うまく A, B, C, D を考えれば 16 類型に分類するのは可能でしょうか？それともこのデータでは不可能で、このデータとは違う 3 種類のデータを考えれば 16 類型に分類するのは可能でしょうか？または、どんなデータでも不可能なことなのでしょうか？

このデータで可能ならば、A, B, C, D を、違うデータで可能ならば、そうなるデータの例を、不可能ならその理由を答えてください。

これを一般化して、都道府県のデータに限らず、充分大きい個数のものを n 個の評価で 2^n 類型に分類できる条件はなんであるか考えてみてください。

	X	Y	Z		X	Y	Z		X	Y	Z
北海道	83457	5506	700	石川	4186	1170	46	岡山	7010	1945	73
青森	9645	1373	181	福井	4190	806	36	広島	8480	2861	89
岩手	15279	1330	156	山梨	4201	863	59	山口	6114	1451	57
宮城	6862	2348	132	長野	13105	2152	159	徳島	4147	785	58
秋田	11636	1086	110	岐阜	9768	2081	79	香川	1862	996	56
山形	6652	1169	124	静岡	7255	3765	166	愛媛	5678	1431	103
福島	13783	2029	151	愛知	5116	7411	171	高知	7105	764	85
茨城	6096	2970	254	三重	5762	1855	93	福岡	4846	5072	141
栃木	6408	2008	141	滋賀	3767	1411	41	佐賀	2440	850	79
群馬	6362	2008	113	京都	4613	2636	39	長崎	4105	1427	111
埼玉	3768	7195	126	大阪	1899	8865	35	熊本	7268	1817	150
千葉	5082	6216	231	兵庫	8396	5588	100	大分	5100	1197	93
東京	2104	13159	39	奈良	3691	1401	32	宮崎	6795	1135	162
神奈川	2416	9048	53	和歌山	4726	1002	63	鹿児島	9044	1706	188
新潟	10364	2374	179	鳥取	3507	589	44	沖縄	2276	1393	67
富山	2046	1093	53	島根	6708	717	48				

出典総務省統計局刊行, 総務省統計研修所編集「日本の統計 2013」

解説

今回の問題は２つの考えるべき点が有りました。１つは、２種類のデータを使い問題のような評価で分類した場合、データの問題では無く、３つの評価では絶対に８種類に分類することは不可能であるということに気づくこと。もう１つは、２種類のデータは２つの評価で４つに分けることはできるのに､何故３つの評価で８種類に分けることができないのか、その理由を説明してもらうことでした。

この２つは同じことのように思えますが、微妙に違います。

1. ３つの基準では絶対に８種類に分類することは不可能であること

まず､データの問題では無く、２種類のデータを使い問題のような評価で分類した場合、３つの評価では絶対に８種類に分類することは不可能であることを考えましょう。

最初に考えなくてはいけないのは、Xが平均以上か平均より小さいかという評価は、Xの平均がポイント、つまりXの平均を$\bar{x}$とすると$X=\bar{x}$のところが規準になり、データを２つに分類しているということです。

同様にYの平均を$\bar{y}$とすると$\mathrm{Y}=\bar{y}$のところが規準になり、$AX+BY$がC以上か、Cより小さいも同様に、$AX+BY\geqq C$と$AX+BY<C$に分けられるので、$AX+BY=C$の所が規準になります。

これを境界と言います。境界によって分けられた部分を領域と言います。領域に新しい境界が通ったとき、元の領域は２つの領域に分けられます。

１つ目の考えるべき点は、２つの境界によって４つの領域に分けられている時に、新しい境界を引いても元の４つの領域のうち少なくとも１つの領域は通らないことを思いつき、自分が納得するだけでなく、誰もが同意するように説明することです。それをやってみましょう。

今、$X=\bar{x}$という境界と、$Y=\bar{y}$という境界によって、４つの領域に分けられている。

もし、 $A\bar{x}+B\bar{y}>C, \mathrm{A}>0, \mathrm{B}>0$であれば、 $X\geqq\bar{x},\ Y\geqq\bar{y}$のとき、

$X-\bar{x}\geqq 0,\ Y-\bar{y}\geqq 0$であるから、

$A(X-\bar{x})\geqq 0, B(Y-\bar{y})\geqq 0$であるので、$A(X-\bar{x})+B(Y-\bar{y})\geqq 0$

$$AX-A\bar{x}+BY-B\bar{y}\geqq 0$$

$$AX+BY\geqq A\bar{x}+B\bar{y}>C$$

つまり、３つ目の境界$AX+BY=C$は領域 $X\geqq\bar{x}, Y\geqq\bar{y}$ を通らない。

他の場合も同様に考えると、すべての場合で３つ目の境界 $AX+BY=C$ は４つの領域のうち少なくとも１つを通らない。

大阪教育大学付属天王寺中学校の中島君はこの方法で n 種類のデータの場合まで考えてくれました。

これはグラフで説明することもできます。

$\mathrm{AX}+\mathrm{BY}=\mathrm{C}$は

・$\mathrm{B}\neq 0$のとき、

$$\mathrm{Y}=-\frac{A}{B}X+\frac{C}{B}$$

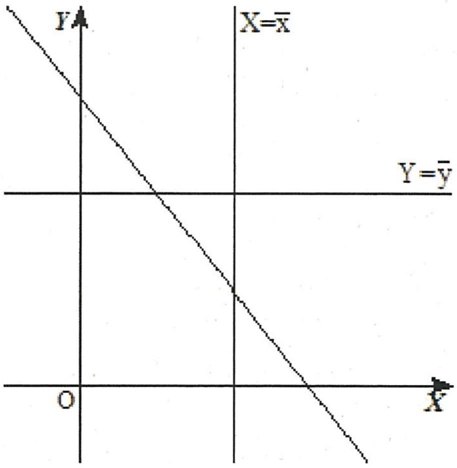
と変形でき直線である。

・$\mathrm{B}=0$のとき、$\mathrm{AX}=\mathrm{C}$となるが、

さらに$\mathrm{A}=0$であると、$0=\mathrm{C}$となり、X, Yのデータに関係が無くなってしまう。

そこで$\mathrm{A}\neq 0$と考えると、

$$\mathrm{X}=\frac{C}{A}$$



グラフ 1

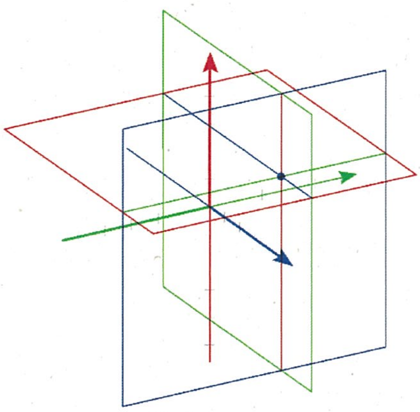
と変形できる。
これも直線である。

したがって、$AX + BY = C$ を付け加えるということは、境界$X = \bar{x}, Y = \bar{y}$によって４つの領域に分けられているところに左のグラフ１のように新しい直線を引くことである。
ところが、どのように直線を引いても３つの領域しか通らない。
つまり、４つのうち少なくとも１つの領域を通らない。

このようにグラフで考えてくれた解答がほとんどでした。

2.　３つの以上のデータの場合

式の大小で考えた解答はデータが３つ以上に増えても同じようにできますが、グラフで考えた解答はデータが３つだと３次元のグラフを、データが４つだと４次元のグラフを考えなくてはいけません。


グラフ 2

３つのデータを考える３次元の場合、$X = \bar{x}$, $Y = \bar{y}$, $Z = \bar{z}$ や $AX + BY + CZ = D$ という境界は平面になります。
左のグラフ２は $X = \bar{x}$, $Y = \bar{y}$, $Z = \bar{z}$ の３つの境界(３つの平面)によって３次元空間が８種類に分けられる様子を表しています。

この状態にもう１つ平面を付け加えても少なくとも１つの領域を通らないので、１６個の領域に分割することは不可能です。

とはいえ、３次元空間の状態を２次元の紙の上に表すのはわかりにくいですね。
ましてや４次元空間の状態や５次元空間の状態を考えるのは相当訓練をしないと難しいでしょう。
しかし、２次元平面、３次元空間で考えたことの類推で、４次元以上のn次元の空間でも同じように、n個の境界までは前の領域のすべてを通り、領域の数を２倍にしていくが、$n + 1$個目の境界は少なくとも１つの領域を通らなくなるだろうと予想できます。

式の大小で考えれば、$n + 1$ 個目の境界が少なくとも１つの領域を通らないことを示すことはできます。

3.　n個のデータは問題のような $n + 1$個の境界では絶対に 2^{n+1}個の領域に分けることは不可能である理由

２つ目の考えるべき点は、何故データの種類以下の数の境界までは領域の数を２倍に増やすことができるのにデータの数を超えた数の境界では領域を２倍に増やすことができな

いのかです。
グラフの場合で考えたことが参考になります。

2次元(データが2種類)の場合は、$X=\bar{x}, Y=\bar{y}$ という境界の共通部分は、1つの点$(\bar{x},\bar{y})$(交点)になり、3つめの境界でこの点を2つに分けるのは不可能です。
ところが、3次元(データが3つ)の場合は、$X=\bar{x}, Y=\bar{y}$ という境界の共通部分は Z が何でも良いので、空間内の直線(これを交線といいます)になります。直線は 3つめの境界で2つに分けることが可能です。

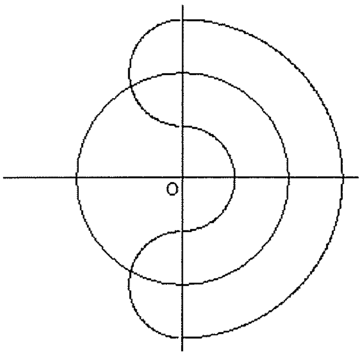
すべての境界の共通部分があれば、その周りにはすべての領域があります。
それまでにあるすべての境界の共通部分を新しい境界で2つに分けることができれば、新しい境界はすべての領域を2つに分けることができますから、領域は2倍の数になります。

つまり、
それまでにあるすべての境界の共通部分が新しい境界で2つに分けられるうちは領域を2倍に増やすことができる。
それまでにあるすべての境界の共通部分が1点であったり、存在しなければ、新しい境界で領域を2倍に増やすことはできない。
と言えそうです。

実際に、4次元以上のn次元の空間を表す n個のデータ、$(X,Y,Z,\cdots,U)$では、n個の境界 $X=\bar{x}, Y=\bar{y}, Z=\bar{z}, \cdots, U=\bar{u}$ の共通部分が1点 $(\bar{x},\bar{y},\bar{z},\cdots,\bar{u})$ になります。
ここまでは領域の数を2倍にしていき、2^n 個の領域に分かれるが、$n+1$ 個目の境界は少なくとも1つの領域を通らないと言えます。

これは、n個の未知数を決めるためには n個の方程式 (境界)が必要である。という連立方程式の原理と同じです。

4. 問題以外の境界の場合を考える



グラフ 3

境界が円や球や4次元以上の空間でのこれに相当するものになっても上の原理は変わりませんが、n次元の空間での n個の境界の共通部分は2点になります。この2点を2つに分けるように $n+1$個目の境界を作ることは可能なので、$n+1$個の境界で、2^{n+1}個の領域に分けることまでは可能です。しかし、$n+1$個の境界の共通部分は存在しないので、$n+2$個の境界で、2^{n+2}個の領域に分けることは不可能だと言えそうです。
開成中学校の伊佐君はこれに近いアイディアまで考えてくれました。

ところが、さらに複雑な境界を考えたときグラフ3のように2次元平面で 4個の境界を使って、$2^4=16$個の領域に分けることが可能です。
この場合も、「領域に新しい境界が通ったとき、元の領域は2つの領域に分けられる。」から、今あるすべての領域を通る境界を引けばよいと言えますが、どの次元でも引くことが可能かどうかはまだ考えてみる余地があります。
みなさんの健闘を期待します。