

(JAPLA 2010/5/22)

順序関係のある分割表に対する順位連関係数(rank association coef)

統計数理研究所(名誉教授) 鈴木義一郎

下の表は、統計数理研究所が1998年に行った『日本人の国民性調査』の質問で「生活が豊かになるか、それとも貧しくなると思うか」についての予想を問うた回答を、学歴別に集計した結果である。豊かになるか否かの見通しと学歴との間の関連性の程度を示すような《尺度》について考えてみよう。

学歴	小学校卒	中学校卒	高校卒	大学卒
豊かになる	16(24.1)	32(11.8)	86(13.5)	64(18.0)
変わらない	18(27.3)	86(13.5)	163(25.6)	99(27.8)
貧しくなる	22(33.3)	134(49.5)	343(53.8)	169(47.5)
その他	0(0.0)	1(0.4)	2(0.3)	1(0.3)
D. K.	10(15.2)	18(6.6)	43(6.8)	23(6.4)
計	66(100)	271(100)	637(100)	356(100)

与えられたデータから、その他とD・K・(無回答)を除いたものをGに入力する。

G=:16 32 86 64,18 86 163 99,:22 134 343 169		学歴別国民性調査のデータ	
G]p=:(<0 0){G	+/,q]r=:p*+/,q
16 32 86 64	16	994	15904
18 86 163 99]q=: (1-4){."1}.G	rは、生活の分類基準で第1カテゴリーが学歴の分類基準で第1カテゴリーからみて順位の逆転していない観測対の総数。	
22 134 343 169	86 163 99 134 343 169		

div=:4 :0 p=.(<<:'a b'=.x){y p;(b-# :y){."1 a}.y)	1 1 div G 16 86 163 99 134 343 169	pos=:4 :0 p=.(<<:'a b'=.x){y p*+/, (b-# :y){."1 a}.y)
---	--	---

]p11=:1 1 pos G 15904]p12=:1 2 pos G 24768]p13=:1 3 pos G 23048]p21=:2 1 pos G 11628
]p22=:2 2 pos G 44032]p23=:2 3 pos G 27547	+ / p11, p12, p13, p21, p22, p23 146927	

<pre> racp=:3 :0 a=.r=.0 ['c d'=.<:\$y while.a<c do.a=.a+1+b=.0 while.b<d do.r=r+(a,b=.b+1)pos y end. end.) </pre>	racp G 146927	2組の分類基準で順位が逆転してない観測対の総数。
	racp .G 149385	2組の分類基準で順位が逆転している観測対の総数。
	N. B. 順位連関係数 rac=:[:(-/%+)] [[:racp" _1],: .]	.G 22 134 343 169 18 86 163 99 16 32 86 64
	rac G _0.00829531	

]G5=::(5*{"1 G)0} :G 80 32 86 64 90 86 163 99 110 134 343 169	rac G5 0.102805 modify=:[:(:%1{:)],+/ rac_m=:[:rac modify	rac_m G 0.122717 rac_m G5 0.122717
modify G 0.285714 0.126984 0.14527 0.192771 0.321429 0.34127 0.275338 0.298193 0.392857 0.531746 0.579392 0.509036	modify G5 0.285714 0.126984 0.14527 0.192771 0.321429 0.34127 0.275338 0.298193 0.392857 0.531746 0.579392 0.509036	
GとG5の順位連関係数が異なるというのは不都合である。これは回答者数がカテゴリー間で不揃いのまま係数を算出したからである。「modify」で基準化してからならO. K. 高学歴になるほど「貧しくなる」との予想が多くなるという傾向が認められる。		

A=:35 35 38 32 29 102,53 62 62 65 85 40,:91 104 137 166 120 55			
A 35 35 38 32 29 102 53 62 62 65 85 40 91 104 137 166 120 55]A5=:5{"1 A 35 35 38 32 29 53 62 62 65 85 91 104 137 166 120]A4=:4{"1 A 35 35 38 32 53 62 62 65 91 104 137 166	
同じデータを年齢階級別に集計したもの。Aは20歳代・30歳代・40歳代・50歳代・60歳代・70歳以上の6階級。A5は70歳以上を除いたもの。A4は20~50歳代の4階級			
rac_m A _0.16798	rac_m A5 0.0643254	rac_m A4 0.131946	70歳以上の回答パターンが特に異常(楽天的)なため、負のassociationになっている。

SRCは<http://www.japla.ijfo> のworkshop may 2010からDLできる