量的変量を質的変量に変換して分散分析の独立変数として用いる場合の問題点

富　原　一　哉

分散分析（analysis of variance; ANOVA）は，３つ以上の平均値間の差の統計的検定手法として，最も利用頻度の高い検定法である。この検定では，各独立変数はその「要因」を構成する各「水準」へと割り当てられる。この分散分析における「水準」は本来名義尺度であり，分類上の意味での同一性しか持たない。しかしながら，実際には，順序尺度，間隔尺度，比例尺度のデータを便宜的に名義尺度に変換することによって，これを分散分析の独立変数として用いることが多く認められる。このように，量的変量を質的変量として分散分析の独立変数として用いた場合には，いくつかの問題が生じやすい。本稿では，これらの問題によって心理統計の初学者が陥りやすい統計処理上の誤りを整理して示すことにより，統計的検定の運用における注意を喚起したいと思う。

１．　情報量の低減

通常，量的変量は質的変量よりもそれぞれの変数が持つ情報が多い。したがって，量的変量を質的変量に変換した場合，そのことにより独立変数と従属変数との関係性を示すための重要な情報が失われる可能性がある。例えば，図１のように独立変数Ｘと従属変数Ｙとの間に，逆Ｕ字型の２次関数的関係が認められた場合，これを独立変数Ｘの平均値で「高群」「低群」に分割すると，Ｙの平均値は両群間に差がなく，「独立変数Ｘは従属変数Ｙに影響を与えない」という誤った結論を導く可能性がある。もちろん，独立変数Ｘを「高群」「中群」「低群」と３群に分ければ，この場合は問題が解決するが，

どのように分割するのが適切であるかは，独立変数Xと従属変数Yの関係性に依存する。例えば，図2に示されているような変数の場合は，独立変数をどこで分割すれば適切といえるのか，非常に判断が難しい。したがって，独立変数と従属変数の関係性を明確に示すことができるよう，本来の量的変量のままで両者の関係性を記述した方が，正確で詳細な説明が可能である[1)]。

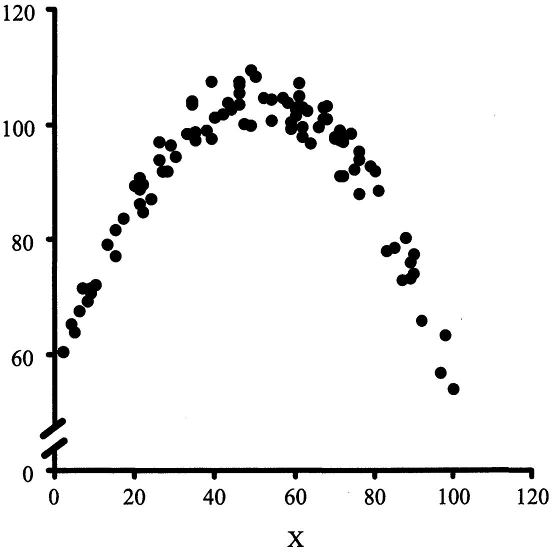

図1　2次関数的関係にある2変数

もちろん，量的変量を質的変量に変換することには，単純化することにより変数間の関係性が理解しやすくなるというメリットはある。例えば，独立変数と従属変数とが単純な1次関数的関係を持っていたならば，両変数間の相関係数が非常に低くとも，十分な標本数を用いることによって，平均値間に明確な統計的有意差を導くことができる。例えば，表1と図3に示されたデータの例では，両変数の相関係数[2)]は $r = 0.163$ ($p > .05$) にすぎないが，独立変数を「高群」「低群」の2つに分割した分散分析においては群間に有意な差を認めることができる ($F(1/98) = 4.68$, $p < .05$)[3)]。このような単純化がはたして適切であるか否かは，対象となる現象の性質や研究の目的によって異なってくるだろう。したがって，その適切性を判断するためにも，変数を変換する前に，まず本来の形でその関係性を十分に吟味しておく必要があると言える。

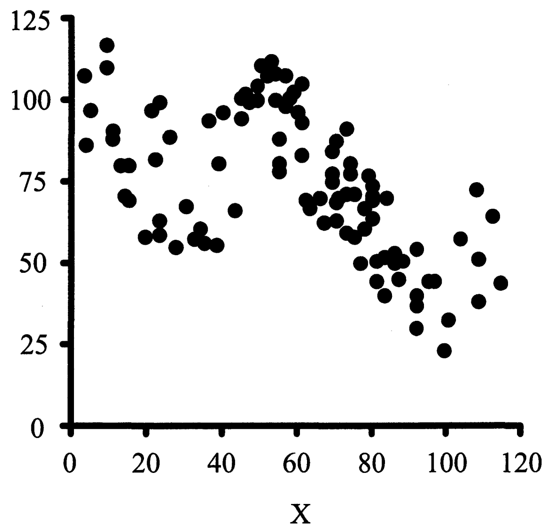

図2　複雑な関数関係をもつ2変数

さらに言えば，分散分析では群間の平均値の差のみを検定しているため，

表1　弱い相関関係を持つ2変数の例

	X	Y		X	Y
	100	77		50	54
	99	99		50	11
	99	29		49	45
	98	33		46	39
	97	23		46	96
	96	97		45	28
	95	45		45	86
	95	55		45	43
	93	49		45	25
	92	61		45	44
	91	100		41	16
	91	69		39	50
	89	72		38	32
	89	56		38	44
	89	25		38	59
	88	70		38	84
	88	75		36	19
	88	70		34	44
	87	62		33	35
	85	45		32	27
	84	10		30	40
	83	74		30	56
高群	83	53	低群	29	51
	82	89		29	100
	78	19		28	25
	78	69		26	45
	77	45		25	25
	76	88		25	42
	74	60		24	81
	73	92		24	60
	71	58		22	24
	71	98		22	37
	70	83		21	27
	69	58		21	50
	65	14		21	61
	65	26		20	31
	63	72		20	71
	62	82		20	90
	62	83		19	62
	61	97		16	71
	60	84		15	66
	59	64		15	81
	58	28		14	50
	56	28		14	54
	54	76		13	90
	54	25		12	63
	54	78		11	16
	52	48		10	50
	52	22		10	46
	52	66		10	37

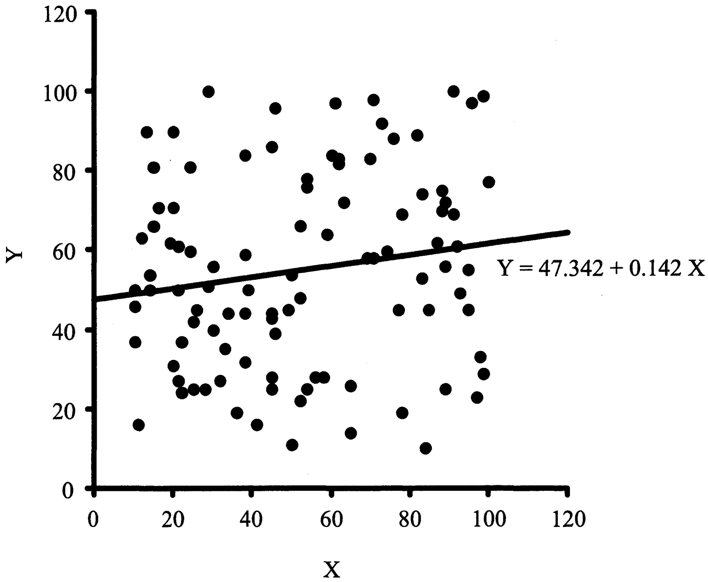

図３　表１のデータに基づく弱い相関関係を持つ２変数の散布図

独立変数と従属変数との間に１次関数的関係があるのか，２次関数的関係があるのか，あるいはまたさらに高次への関数関係を想定すべきかといった点については明確にできない。このような分析には，直交多項式を用いた２次以上の式への当てはめを行う非線形回帰分析（Kirk，1982）が適しており，多くの場合これは分散分析による単純化された分析よりも有益な情報をもたらすものと思われる。

２．　要因の交絡

量的変量を質的変量に変換して分散分析の独立変数として用いる場合の，もっと重要で致命的な問題は，２要因以上の分散分析において，複数の独立変数間に相関関係が認められた場合に起こる。例えば，表２の例では，２つの独立変数ＸとＹが，それぞれの平均値を境に「高群」と「低群」に分割されている。この群分けに基づいて，従属変数Ｚについての２×２の級間２要因分散分析を行うと，表３の分散分析表に示された結果が得られる。つまり，要因Ｘ（$F(1/16) = 83.36$，$p < .0001$）と要因Ｙ（$F(1/16) = 17.15$，$p < .001$）のどちらも主効果が有意であり，「Ｘ高群」は「Ｘ低群」よりも，「Ｙ高群」は「Ｙ低群」よりもＺの値が高いと結論される。しかしながら，実はこの独立変数ＸとＹの間には強い相関関係（$r = 0.645$，$p < .01$）があり，要因Ｙで認められた主効果は，これによって歪められた結果有意となってしまっているのである。

これを簡単に理解するために表３のＸとＹとを２次元上で表した図４を参

表2　2つの独立変数間に相関関係のあるデータの例

		X（要因X）	Y（要因Y）	Z（従属変数）
X1（Low）	Y1（Low）	24	12	12
		26	22	16
		31	23	18
		38	26	23
		39	25	38
	Y2（High）	31	39	21
		36	33	31
		51	37	41
		56	32	46
		57	44	46
X2（High）	Y1（Low）	58	21	42
		64	28	58
		66	26	62
		68	27	64
		73	30	65
	Y2（High）	74	39	70
		82	41	73
		91	34	78
		93	48	79
		99	49	89

表3　表2のデータに基づく分散分析

factor	SS	df	MS	F
X	7527.20	1	7527.20	83.36***
Y	1548.80	1	1548.80	17.15**
X ＊ Y	20.00	1	20.00	0.22
error	1444.80	16	90.30	
total	10540.80	19		

***$p < .0001$，**$p < .001$

照されたい。ここでは「X高Y高群」「X高Y低群」「X低Y高群」「X低Y低群」の各群は，それぞれa，b，c，dの線で囲まれた部分に分布している。このとき，aの「X高Y高群」とbの「X高Y低群」は，ともに「X高群」でありながら，それぞれの群内の変数Xの平均値 $\overline{X}a$ = 87.8と $\overline{X}b$ = 65.8は有意に異なっている（t (8)=4.37，p ＜ .01)。したがって，「X高Y高群」と「X高Y低群」は変数Yについて異なっていると同時に，変数Xについても異なっているのである。同じことは，「X低Y高群」と「X低Y低群」の間ではもちろんのこと，「X高Y高群」と「X低Y高群」の間，「X高Y低群」と「X低Y低群」との間でも言える。したがって，要因Xと要因Yは完全に交絡しており，このデータに基づいて分散分析を行っても，要因Xと要因Yのそれぞれの主効果は，変数Xの相違によって起こっているものなのか，あるいは変数Yの相違によって起こっているものなのか，全く確定できないことになる[4)]。

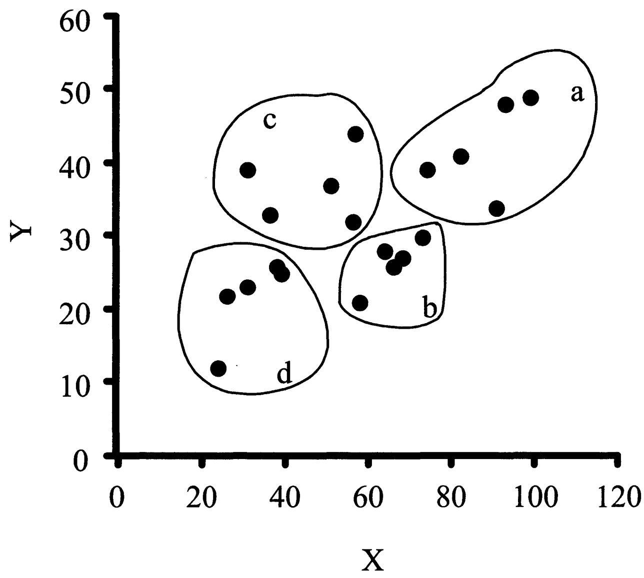

図4　要因が交絡した群分けの例

一方，これらの3変数間の偏相関は，変数Xと変数Yのどちらが変数Zに本質的に影響を及ぼしていたのかを確定するための指標となりうる。変数X，Y，Zの偏相関係数を計算すると，XとZ間ではr = 0.974，YとZ間ではr = -0.022となり，分散分析によって示されたYのZに対する主効果は，Xによって媒介された疑似相関による部分が大きいことが明らかとなる。

なお，同一のデータでX－Y間に強い相関がない場合を表4に示した。個々のデータが等しいので分散分析では表3と全く同じ結果が得られる。しかしながら，偏相関係数を計算すると，XとZ間ではr = 0.981，YとZ間では

表4　2つの独立変数間に相関関係のないデータの例

		X（要因X）	Y（要因Y）	Z（従属変数）
X1（Low）	Y1（Low）	24	21	12
		39	23	16
		36	26	18
		38	22	23
		56	27	38
	Y2（High）	31	37	21
		31	39	31
		26	32	41
		51	44	46
		57	49	46
X2（High）	Y1（Low）	64	12	42
		74	28	58
		73	26	62
		66	25	64
		99	30	65
	Y2（High）	68	39	70
		82	41	73
		58	48	78
		91	34	79
		93	33	89

$r = 0.581$となり，XとYの両変数とも変数Zに対して強い影響を及ぼしていることが分かる。このように，独立変数間に相関関係が認められない場合は，分散分析でも偏相関による分析でも結果に大きな違いはない。しかしながら，先の例のように独立変数間に相関関係が認められる場合は，誤った結論を導く可能性が高いので，量的変量を質的変量の独立変数に変換して分散分析の独立変数として用いるべきではないと言える。

当然のことであるが，このように複数の量的変量を独立変数として用いる場合は，あえてそれを質的変量に変換することなく，はじめから重回帰分析（大塚，1988）や共分散分析（渡部，1988），あるいは共分散構造分析（豊

田，1998）等を適用した方が適切であろう。その場合においても，線形性や変数間の相関の問題は時として重要な影響を及ぼしうるので，あらかじめ取り扱う変数の特徴について十分に吟味することが必要であると言える。

注

1）このような場合は，相関比（吉田，1990）を関係性の指標として用いる。

2）本論文では，相関係数は全てピアソンの積率相関係数を用いている。

3）ｔ検定によっても同じ結果が得られる。

4）このような歪みは，中央値を基準に群分けを行った場合でも等しく起こりうる点に注意されたい。

引用文献

Kirk, R. E. 1982 *Experimental design: Procedures for the behavioral sciences. (2nd ed.)* Monterey: Brooks/Cole.

大塚雄作 1988　重回帰分析　渡部洋（編）心理教育のための多変量解析入門（基礎編）3章　福村出版　57-77.

豊田秀樹　1998　共分散構造分析 入門編－構造方程式モデリング　統計ライブラリー　朝倉書店.

渡部洋 1988　共分散分析　渡部洋（編）心理教育のための多変量解析入門（基礎編）6章　福村出版　121-132.

吉田寿夫　1990　2つの変数の関係を分析する方法　森敏昭・吉田寿夫（編）心理学のためのデータ解析テクニカルブック　第5章　北大路書房　217-259.