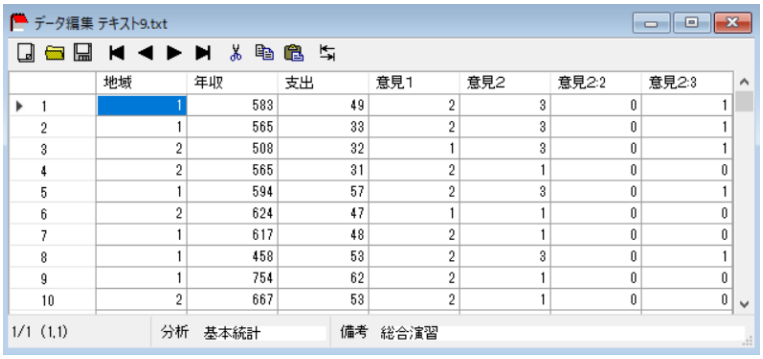
分析間の関係

多くの検定や分析を扱っていると、個別に理論を学んでいるときには見逃しがちな分析間の同一性が見えてくることがある。それは特に分析の中に 2 分類（例えば 0/1）の変数が含まれるときに多く見られる。ここではその分析間の類似性についてまとめておく。

１．相関係数・ｔ検定・数量化Ⅰ類・χ^2検定

相関係数と単回帰分析の検定でよく知られていることは、相関係数の検定と単回帰式の傾きの係数の検定が同一であることである。また、単回帰式の傾きの係数の検定が単回帰式の有効性の検定と同一であることも周知のことがらである。しかし、説明変数が 2 つ以上の重回帰分析の場合は、目的変数と各説明変数の相関係数の検定と偏回帰係数の検定は異なり、重回帰式の有効性の検定もこれらとは異なる。このようなよく知られた話も含めて、分析間の関係について例を用いて述べて行く。

図 1 にこの節で利用するデータを示す。最後の 2 列は 3 分類の意見 2 を 0/1 データに変換し、その 1 列目を取り除いたものである。



	地域	年収	支出	意見1	意見2	意見2-2	意見2-3
1	1	583	49	2	3	0	1
2	1	565	33	2	3	0	1
3	2	508	32	1	3	0	1
4	2	565	31	2	1	0	0
5	1	594	57	2	3	0	1
6	2	624	47	1	1	0	0
7	1	617	48	2	1	0	0
8	1	458	53	2	3	0	1
9	1	754	62	2	1	0	0
10	2	667	53	2	1	0	0

1/1 (1,1)

図 1　利用するデータ

まず地域で分けた年収の差の t 検定、地域を量的データとみなした相関係数の検定、年収を目的変数、地域を説明変数とした数量化 I 類による分析を実行する。それぞれの分析結果を図 2a、図 2b、図 2c に示す。

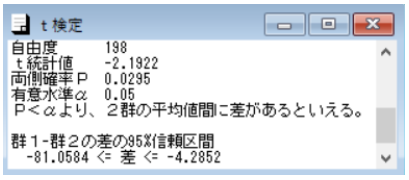


図 2a　t 検定結果

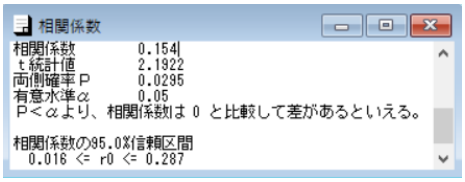


図 2b　相関係数結果

	カテゴリウェイ	重回帰ウェイ	基準化ウェイ
地域:1	608.1489	0.0000	-22.6161
地域:2	650.8208	42.6718	20.0558
定数項	0.0000	608.1489	630.7650
重相関R	0.154	調整済R	0.137
寄与率R^2	0.024	調整済R^2	0.019
有効性F値	4.806	自由度	1,198
p値	0.0295		

図 2c　数量化 I 類結果

これらの検定確率値はすべて p=0.0295 である。数量化 I 類の分析は F 検定値であるが、これは t 検定値を 2 乗したものである。数量化 I 類は重回帰分析と同等であるので、相関係数の検定が回帰分析と同等であることから、結果は当然である。

以上は 2 分類の質的データを用いたものであったが、3 分類以上では、t 検定に相当する検定が 1 元配置分散分析、相関係数が分類を 0/1 データに変えた正準相関分析、数量化 I 類は説明変数が 3 分類以上の質的変数になるだけである。

まず、意見 2 と年収を選んで 1 元配置分散分析を行った結果を図 3a に、年収を目的変数に、意見 2 を説明変数に選んで数量化 I 類を実行した結果を図 3b に示す。

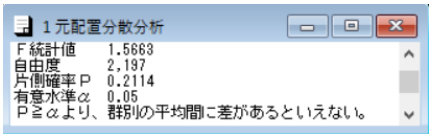


図 3a　1 元配置分散分析結果

図 3b　数量化 I 類結果

検定確率で見ると p=0.2114 と 1 元配置分散分析と数量化 I 類は分類が増えても同一である。また相関係数については、図 3b と図 4 に示すように、数量化 I 類では重相関係数の r=0.125、正準相関分析では年収と意見 2 :2、意見 2 :3 を 2 つの正準変数に選んだ場合の正準相関係数の r=0.125 に拡張される。

図 4　正準相関分析結果

次に 2 分類の地域と意見 1 について、相関係数と χ^2 検定を比較する。図 5a と図 5b にそれぞれの検定結果を示す。但し、χ^2 検定ではイェーツ補正を行っていない。

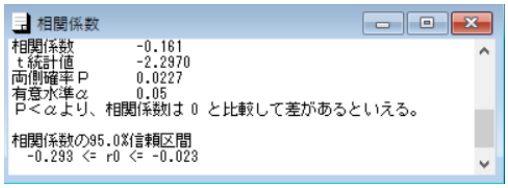


図 5a　相関係数結果

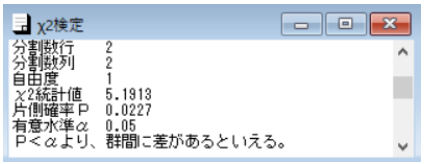
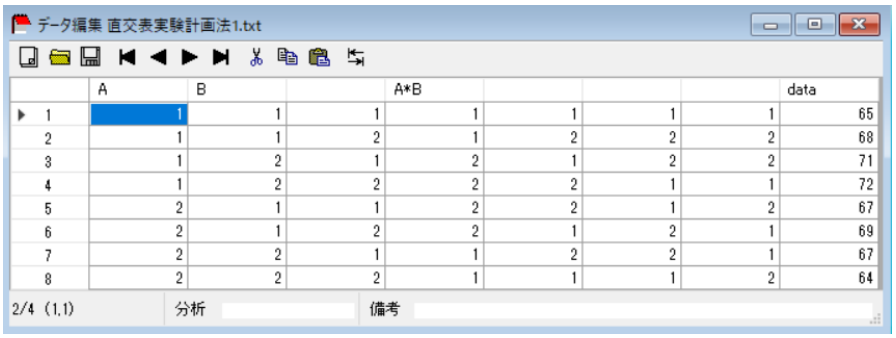


図 5b　χ^2 検定結果

2 分類のデータでは、相関係数の検定と χ^2 検定は検定確率 p=0.0227 で同等である。しかし、3 分類以上のデータが含まれる場合、異なった結果になる。3 分類以上を量的データとして扱った場合、分類間の距離をどうとらえるかという基準が明確でないため、χ^2 検定と同一にならない。

２．２元配置分散分析・直交表分散分析・数量化Ｉ類

前節で１元配置分散分析と数量化Ｉ類の有効性の検定が同一であることを述べたが、２元配置分散分析は他の検定とどのような関係があるのだろうか。ここで、重要な役割を演じるのは直交表分散分析である。直交表分散分析は実験に影響を与えるすべての要因の条件の数を、直交表と呼ばれる組み合わせ表でそろえる分析である。この分析により、最小限の実験回数で交互作用を持つデータの差の検定が実施できる。図 6 に 2 分類の直交表実験計画法 L8(2^7)のデータを示す。



	A	B		A*B			data
1	1	1	1	1	1	1	65
2	1	1	2	1	2	2	68
3	1	2	1	2	1	2	71
4	1	2	2	2	2	1	72
5	2	1	1	2	2	1	67
6	2	1	2	2	1	2	69
7	2	2	1	1	2	2	67
8	2	2	2	1	1	1	64

図 6　直交表実験計画法のデータ

要因 A, B による data の違いを調べた 2 元配置分散分析の結果と直交表分散分析の結果を図 7a と図 7b に示す。C.Analysis の直交表分散分析は図 6 の変数をすべて選択して実行する。

	平方和	自由度	不偏分散	F値	確率値
全変動	52.875	7			
A水準間	10.125	1	10.125	3.5217	0.1338
B水準間	3.125	1	3.125	1.0870	0.3560
相互作用間	28.125	1	28.125	9.7826	0.0353
水準内	11.500	4	2.875		

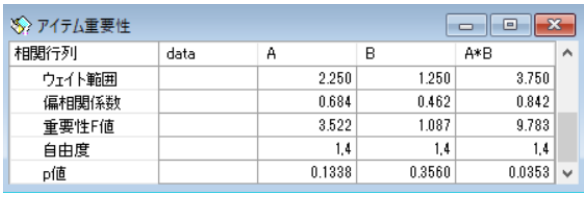
図 7a　２元配置分散分析結果

要因	平方和S	自由度	平均平方V	F値	P値
A	10.125	1	10.125	3.522	0.1338
B	3.125	1	3.125	1.087	0.3560
A*B	28.125	1	28.125	9.783	0.0353
誤差	11.500	4	2.875		
Total	52.875	7			

図 7b　直交表分散分析結果

図 7a と図 7b の確率値はすべて同じである。2 元配置分散分析はデータの一部が欠損しても、データが増えても、何らかの結果を与えるが、直交表分散分析では指定されたデータの形式が必要である。

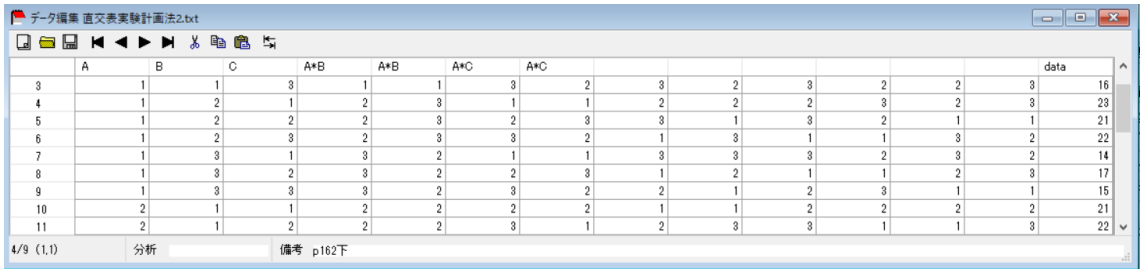
この分散分析を数量化Ｉ類と比較してみよう。目的変数に data、説明変数に A と B と A*B を指定した分析結果を図 8 に示す。2 元配置分散分析及び直交表分散分析の要因による検定結果と数量化Ｉ類のアイテムごとの検定結果は確率が一致している。しかし、データを一部欠損にすると、２元配置分散分析の結果と数量化Ｉ類の検定結果は異なる。


アイテム重要性

相関行列	data	A	B	A*B
ウェイト範囲		2.250	1.250	3.750
偏相関係数		0.684	0.462	0.842
重要性F値		3.522	1.087	9.783
自由度		1,4	1,4	1,4
p値		0.1338	0.3560	0.0353

図 8　数量化 I 類結果

次に直交表分散分析で交互作用が複数列で表されるような場合について検討する。図 9 に 3 分類の直交表実験計画法 L27(3^13)のデータを示す。


データ編集 直交表実験計画法2.txt

	A	B	C	A*B	A*B	A*C	A*C							data
3	1	1	3	1	1	3	2	3	2	3	2	2	3	16
4	1	2	1	2	3	1	1	2	2	2	3	2	3	23
5	1	2	2	2	3	2	3	3	1	3	2	1	1	21
6	1	2	3	2	3	3	2	1	3	1	1	3	2	22
7	1	3	1	3	2	1	1	3	3	3	2	3	2	14
8	1	3	2	3	2	2	3	1	2	1	1	2	3	17
9	1	3	3	3	2	3	2	2	1	2	3	1	1	15
10	2	1	1	2	2	2	2	1	1	2	2	2	2	21
11	2	1	2	2	2	3	1	2	3	3	1	1	3	22

4/9 (1,1)　分析　備考 p162下

図 9　3 分類直交表実験計画法データ

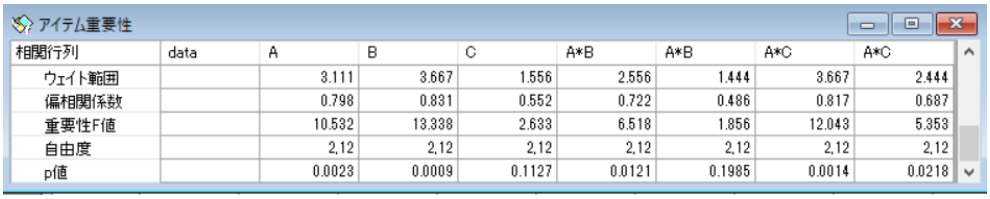
この場合、交互作用は 2 列で与えられる。直交表分散分析の結果を図 10 に示す。


直交表分散分析結果

要因	平方和S	自由度	平均平方V	F値	P値
A	54.222	2	27.111	10.532	0.0023
B	68.667	2	34.333	13.338	0.0009
C	13.556	2	6.778	2.633	0.1127
A*B	43.111	4	10.778	4.187	0.0238
A*C	89.556	4	22.389	8.698	0.0016
誤差	30.889	12	2.574		
Total	300.000	26			

図 10　直交表分散分析結果

このデータで data を目的変数に、変数 A, B, C と 2 つの A*B，2 つの A*C を説明変数に選択して数量化 I 類を実行した結果が図 11 である。


アイテム重要性

相関行列	data	A	B	C	A*B	A*B	A*C	A*C
ウェイト範囲		3.111	3.667	1.556	2.556	1.444	3.667	2.444
偏相関係数		0.798	0.831	0.552	0.722	0.486	0.817	0.687
重要性F値		10.532	13.338	2.633	6.518	1.856	12.043	5.353
自由度		2,12	2,12	2,12	2,12	2,12	2,12	2,12
p値		0.0023	0.0009	0.1127	0.0121	0.1985	0.0014	0.0218

図 11　数量化 I 類結果

A, B, C の結果は直交表分散分析結果と同一である。交互作用 A*B と A*C は 2 つに分かれているが、この 2 つの変数について、結合仮説検定を実施した結果が図 12 である。


重回帰分析

結合仮説の検定（係数は変数名で表します）
結合係数：A*B:2=0,A*B:3=0,A*B:2=0,A*B:3=0
F検定値　4.1871
自由度　4 , 12
確率値　0.0238

重回帰分析

結合仮説の検定（係数は変数名で表します）
結合係数：A*C:2=0,A*C:3=0,A*C:2=0,A*C:3=0
F検定値　8.6978
自由度　4 , 12
確率値　0.0016

図 12　結合仮説検定の結果

この計算にはソフトの制約から、一度データを重回帰分析用に 0/1 データに変換し（数量化 I 類も同じ）、その後 A*B と A*C に関連する部分で結合仮説検定を行った。この結果の

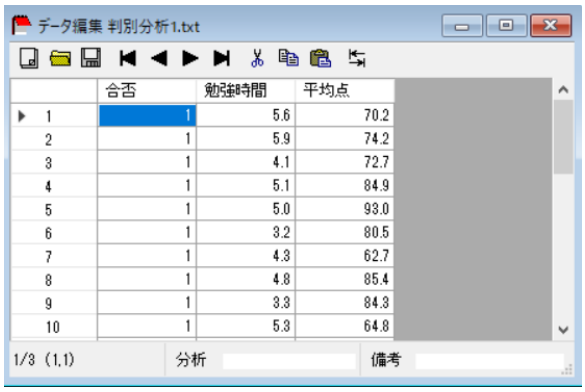
検定確率は直交表分散分析の結果と一致している。

図 9 に与えられたデータで、変数 A, B, A*B, A*B だけを使って、2 元配置分散分析、直交表分散分析、数量化 I 類を行ってもすべての分析が同一の結果を得る。但し、直交表分散分析については、ソフトの性質上、他の変数名は空白にする必要がある。

1 元配置分散分析はどのようなデータでも数量化 I 類の結果と一致するが、交互作用を持つ 2 元配置分散分析は図 9 のようなデータに欠損があれば異なった結果となる。

３．重回帰分析・判別分析

ここでは重回帰分析と判別分析の一致性を議論する。重回帰分析は目的変数自身を推測する分析であり、判別分析は目的変数の分類を判別関数値として与えるものであるので、目的変数が 2 分類のデータである場合、類似性があるものと予想される。図 13 に利用するデータを示す。



	合否	勉強時間	平均点
1	1	5.6	70.2
2	1	5.9	74.2
3	1	4.1	72.7
4	1	5.1	84.9
5	1	5.0	93.0
6	1	3.2	80.5
7	1	4.3	62.7
8	1	4.8	85.4
9	1	3.3	84.3
10	1	5.3	64.8

図 13　判別分析のデータ

合否を目的変数、勉強時間と平均点を説明変数として重回帰分析を行った結果を図 14a に、判別分析を行った結果を図 14b に示す。

偏回帰係数と検定

合否	偏回帰係数	標準化係数	t検定値	自由度	確率値	相関係数	偏相関係数
勉強時間	-0.2369	-0.5484	-4.4589	27	0.0001	-0.613	-0.651
平均点	-0.0212	-0.4767	-3.8765	27	0.0006	-0.551	-0.598
切片	3.9540	0.0000	9.6370	27	0.0000		
R	0.774	R^2	0.599	調整済R	0.755	調整済R^2	0.570

図 14a　重回帰分析結果

判別分析

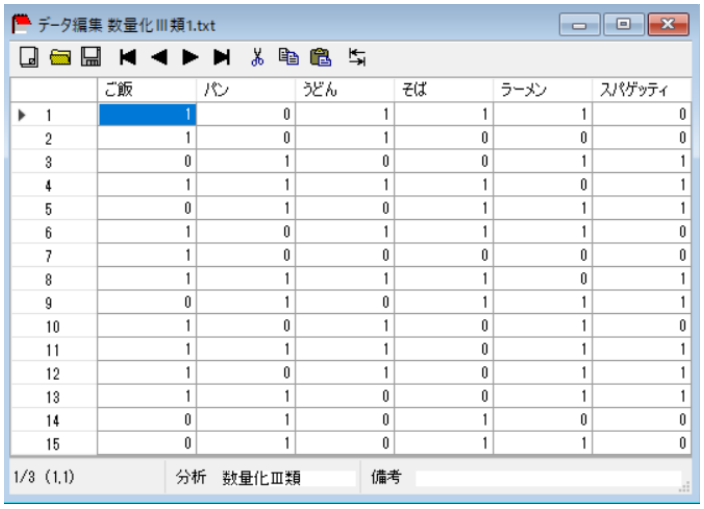
	判別関数	標準化係数	F検定値	自由度	確率
勉強時間	2.246	2.621	19.882	1,27	0.0001
平均点	0.201	2.279	15.027	1,27	0.0006
定数項	-23.019	-0.379			
マハラノビスの距離	5.682				
誤判別確率	1群を2群と	2群を1群と			
理論から	0.117	0.117			
実測から	0.077	0.059			
判別数・確率	1群予測	2群予測	1群予測	2群予測	
1群実測	12	1	0.923	0.077	
2群実測	1	16	0.059	0.941	

図 14b　判別分析結果

図 14a の t 検定値を 2 乗すると図 14b の F 検定値になり、自由度も確率も一致している。

４．数量化Ⅲ類・コレスポンデンス分析

数量化Ⅲ類は 0/1 で与えられたデータから個体や変数の類似性を求める分析であり、コレスポンデンス分析は 2 次元分割表から個体や変数の類似性を求める分析である。そのため、0/1 データを分割表とみなせば、類似性があるように思われる。図 15 に数量化Ⅲ類で用いられるデータを与える。



	ご飯	パン	うどん	そば	ラーメン	スパゲッティ
1	1	0	1	1	1	0
2	1	0	1	0	0	0
3	0	1	0	0	1	1
4	1	1	1	1	0	1
5	0	1	0	1	1	1
6	1	0	1	1	1	0
7	1	0	0	0	0	0
8	1	1	1	1	0	1
9	0	1	0	1	1	1
10	1	0	1	0	1	0
11	1	1	1	0	1	1
12	1	0	1	0	1	1
13	1	1	0	0	1	1
14	0	1	0	1	0	0
15	0	1	0	1	1	0

図 15　数量化Ⅲ類のデータ

このデータを用いて、数量化Ⅲ類を行った固有値と寄与率の結果を図 16a に、カテゴリウェイトの結果を図 16b に示す。

	第1次元	第2次元	第3次元	第4次元	第5次元
固有値	0.351	0.156	0.095	0.060	0.025
相関係数	0.592	0.395	0.308	0.246	0.159
寄与率	0.510	0.227	0.138	0.088	0.037
累積寄与率	0.510	0.737	0.875	0.963	1.000

図 16a　数量化Ⅲ類の固有値・寄与率結果

これに対して、コレスポンデンス分析を行った固有値と寄与率の結果を図 16b に示す。

	群	第1成分	第2成分	第3成分	第4成分	第5成分
固有値		0.351	0.156	0.095	0.060	0.025
相関係数		0.592	0.395	0.308	0.246	0.159
寄与率		0.510	0.227	0.138	0.088	0.037
累積寄与率		0.510	0.737	0.875	0.963	1.000

図 16b　コレスポンデンス分析の固有値・寄与率結果

これを見ると両者は一致している。また、省略しているが、カテゴリウェイトも個体ウェイトも一致する。このように 0/1 データについて、2 つの分析は一致している。