心理データ解析演習　2005/5/11

数量化Ⅱ類

教育認知心理学講座
M1　田村綾菜

数量化理論について

数量化とは

「なんでも数値で表わしてしまうこと」

（大村, 1985）

― どうやって数値を与えるのか？

数量化理論とは

＝「林の数量化理論」

■林知己夫博士が開発したもの。
■理論というよりは、数量化の方法のこと。
■データの種類や解析の目的によって、
Ⅰ類～Ⅳ類の4つの方法に分類される。

林の数量化理論の分類

手法	外的基準	データ	目的	関連する手法
数量化Ⅰ類	量的変数	アイテム・カテゴリー・データ	外的基準の予測	重回帰分析
数量化Ⅱ類	質的変数	アイテム・カテゴリー・データ	外的基準の判別	判別分析
数量化Ⅲ類	なし	アイテム・カテゴリー・データ	変数間の関係の要約と記述	正準相関分析
数量化Ⅳ類	なし	対象間の類似度	対象間の関係の要約と記述	多次元尺度法

（渡部, 1988）

林の数量化理論の分類

手法	外的基準	データ	目的	関連する手法
数量化Ⅰ類	量的変数	アイテム・カテゴリー・データ	外的基準の予測	重回帰分析
数量化Ⅱ類	質的変数	アイテム・カテゴリー・データ	外的基準の判別	判別分析
数量化Ⅲ類	なし	アイテム・カテゴリー・データ	変数間の関係の要約と記述	正準相関分析
数量化Ⅳ類	なし	対象間の類似度	対象間の関係の要約と記述	多次元尺度法

（渡部, 1988）

多変量解析と数量化理論

数量化は、

<u>量的データに対する多変量解析の手法を</u>

<u>質的データにも適用したい</u>

という発想から生まれたもの。

↓

量的データに対する多変量解析の手法と

関連している。

質的データしか扱えないのか？

- 数量化は、もともと質的変数を扱う手法。

- 量的変数でも適当な境界値を設定して
カテゴリー化すれば数量化の対象になりうる。
ex. 年齢を年代としてカテゴリー化する

この場合、量的変数のままでは扱えない
非線形的な関係を扱える、という利点がある。

数量化理論の一般的な表現①

$$\delta_i(j,\ k_j) = \begin{cases} 1 & k_j = K'_j \\ 0 & k_j \neq K'_j \end{cases}$$

iという人がjという項目で、

j分類のK'番目で反応を示すかどうかを表わす。

そうれあればば1、そうでなければ0とする。

このように、0か1しかその値をとらない変数を**ダミー変数**という。

数量化理論の一般的な表現②

各カテゴリカルデータに x_{jk_j} という数値を与えると、

i のもつ数値 α_i は、以下のように表わせる。

$$\alpha_i = \sum_{j=1}^{J} \sum_{kj=1}^{Kj} x_{jk_j} \delta_i(j, kj)$$

ただし、J は項目（アイテム）の総数

Kjは項目jのカテゴリー総数

まとめると・・・

数量化とは、
ダミー変数を用いて、
順序尺度または名義尺度で測定された
質的変数の各カテゴリーに数値を与えて
多変量解析を行う方法のこと。

数量化Ⅱ類について

どんなときに用いるか？

１．外的基準が与えられている。

２．その外的基準が分類（質的変数）である。

例えば･･･

仮釈放の判断を行う　とする。

＜目的＞

裁判に対する態度と受刑中の態度から再犯しないかどうかを予測・判断したい。

仮想データ

数値は渡部(1988)より引用

表1　受刑者の態度と仮釈放の結果

受刑者	再犯	裁判	受刑中
1	1	1	1
2	1	1	2
3	1	1	2
4	1	2	1
5	1	3	1
6	2	1	2
7	2	2	1
8	2	2	2
9	2	2	1
10	2	3	2

再犯
- ある:1
- なし:2

裁判に対する態度
- 不満:1
- どちらでもない:2
- 満足:3

受刑中の態度
- 良:1
- 悪:2

仮想データ

数値は渡部(1988)より引用

表1　受刑者の態度と仮釈放の結果

受刑者	再犯	裁判	受刑中
1	1	1	1
2	1	1	2
3	1	1	2
4	1	2	1
5	1	3	1
6	2	1	2
7	2	2	1
8	2	2	2
9	2	2	1
10	2	3	2

外的基準　　説明変数

再犯
- ある:1
- なし:2

裁判に対する態度
- 不満:1
- どちらでもない:2
- 満足:3

受刑中の態度
- 良:1
- 悪:2

まず

説明変数をダミー変数で表わす

Ex.　受刑者1

$$\delta_1(1,1)=\delta_1(2,1)=1$$

$$\delta_1(1,2)=\delta_1(1,3)=\delta_1(2,2)=0$$

そして

説明変数の影響を和の形で総合した合成変数を作る

Ex.　受刑者1

$$\alpha_1 = 1 \times x_{11} + 0 \times x_{12} + 0 \times x_{13}$$

$$+ 1 \times x_{21} + 0 \times x_{22}$$

$$= x_{11} + x_{21}$$

目標

α の値について、
グループ内でのばらつきを小さく
グループ間でのばらつきを大きく したい。

用いる統計量

グループ内のばらつき(級内平方和)

$$S_w = \sum_{l=1}^{g} \sum_{i=1}^{nl} (\alpha_{li} - \overline{\alpha}_l)^2$$

グループ間のばらつき(級間平方和)

$$S_B = \sum_{l=1}^{g} (\overline{\alpha}_l - \overline{\alpha})^2 n_l$$

全体のばらつき(全平方和)

$$S_T = \sum_{l=1}^{g} \sum_{i=1}^{nl} (\alpha_{li} - \overline{\alpha})^2$$

g:外的基準のグループ数

α_{li}:グループlの第i番目の個体の合成変数値

$\overline{\alpha}_l$:グループlの合成変数値の平均

$\overline{\alpha}$:全個体の合成変数値の平均

N_l:グループlの個体数

$\left(\sum_{l=1}^{g} n_l = N \right)$

目標の確認

S_Wを小さく

S_Bを大きく　したい。

で

一般に、

$$S_T = S_W + S_B$$

が成り立つ。

S_Tが一定ならば、

S_Bを大きくすれば、S_Wは小さくなる。そこで、

$$\eta^2 = S_B / S_T$$

を考える。

η :相関比

$0 \leqq \eta \leqq 1$ の範囲の値をとる。

- $\eta = 1$ であれば、 $S_W = 0$

  つまり、グループ内の合成変数値はすべて等しくなり、
  グループの判別が完全となる。

- $\eta = 0$ であれば、 $S_B = 0$

  つまり、各グループの平均値はすべて等しくなり、
  グループの判別は不可能となる。

つまり

数量化Ⅱ類では、この相関比の2乗 η^2 を最大にするように説明変数の各カテゴリーに数値を与える。

↓

η^2 を x_{jk_j} で偏微分して、それを0とおけばよい。

ややこしいので、計算の過程は省きます。

分析結果①

青木研究室のBlack Boxで処理しました。

アイテム・カテゴリー		カテゴリースコア
裁判	1	−1.18585
	2	1.18585
	3	0
受刑中	1	−0.632456
	2	0.632456

外的基準

再犯あり	－0.632456
再犯なし	0.632456

負の数値は再犯ありの方向に、
正の数値は再犯なしの方向に影響を与えている。

解釈

- 裁判に不満、受刑中の態度が良
  ⇒再犯する傾向
- 裁判はどちらでもない、受刑中の態度が悪
  ⇒再犯しない傾向
- 裁判に満足
  ⇒再犯への影響はみられない
- 受刑中の態度より、裁判に対する態度の方が再犯への影響が強い。

分析結果②

受刑者	サンプルスコア
1	−1.97642
2	−0.39528
3	−0.39528
4	0.39528
5	−0.79057
6	−0.39528
7	0.39528
8	1.97642
9	0.39528
10	0.79057

Ex. 受刑者1

裁判:1、受刑中:1

ー1.18585ー0.632456

＝ー1.97642

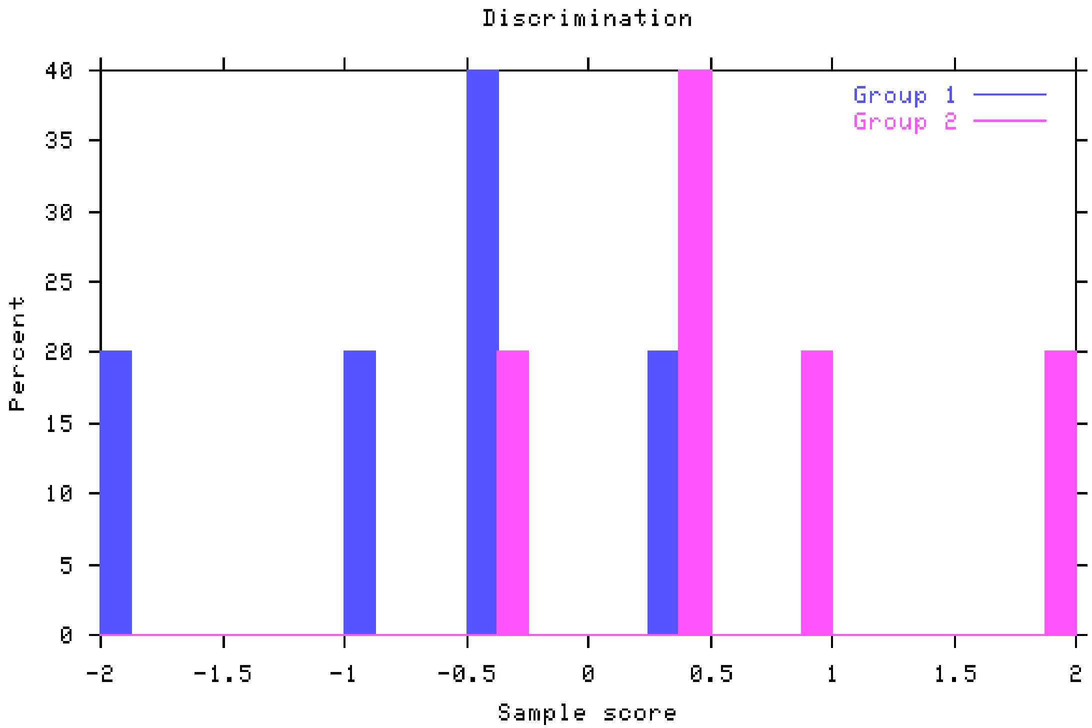
相関比　0.400000

※相関比が小さいので

良い処理がされたとは言えません。

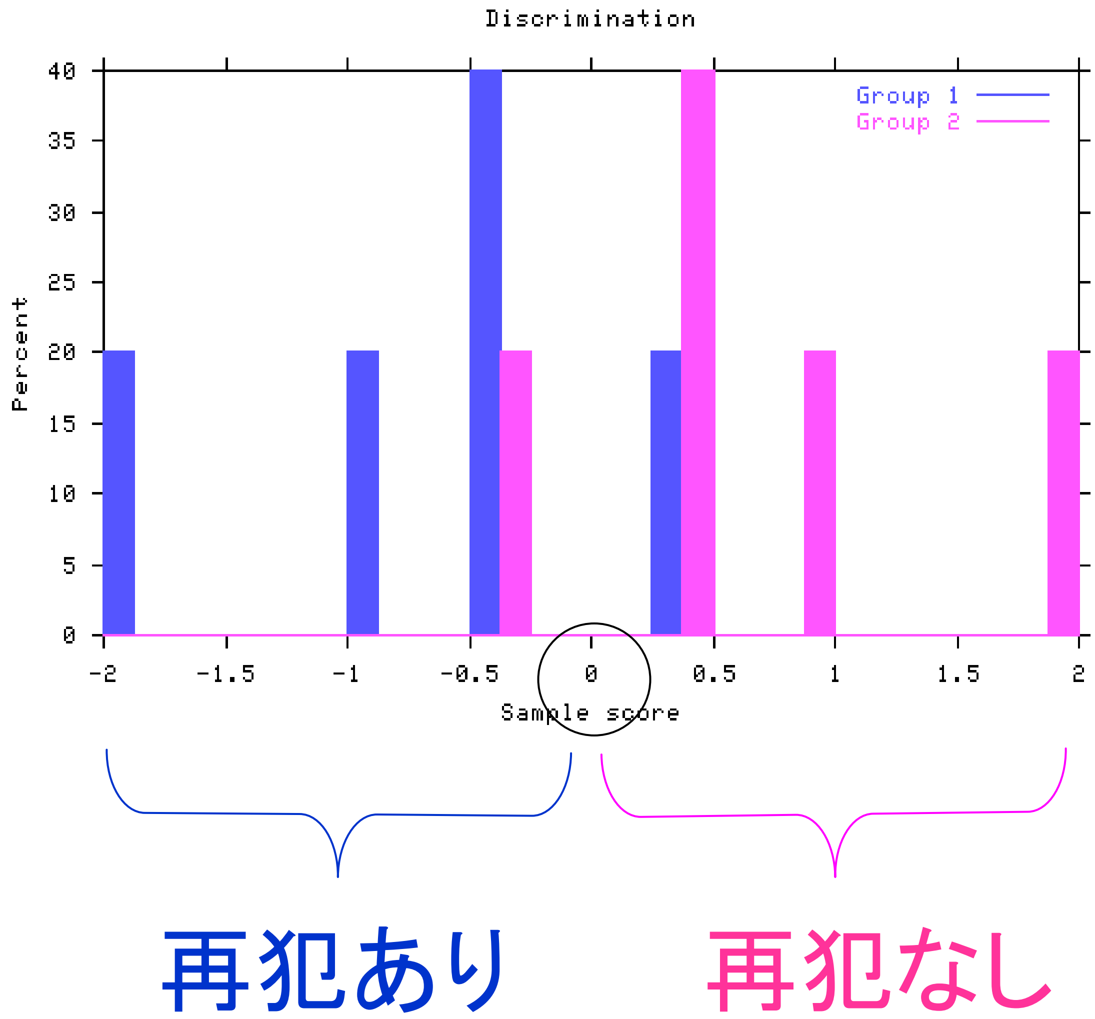
分析結果③

Black Boxで出てくるグラフです。



分析結果③

Black Boxで出てくるグラフです。



カットポイント

0.000000

正判別率　80.00%

判別分析について

判別分析とは

「いくつかの集団や群に属する個人について
多変量データが得られているときに、
各個人が属する集団同士がなるべくよく
区別されるように、多変量データの重みつき
合計点を作ることを目的としたものであり、
そのような合計点をあらわす変数(合成変数)を
判別関数(discriminant function)と呼んでいる。」

(渡部, 1988)

⇒目的は数量化Ⅱ類と同じ。

数量化Ⅱ類との違いは？

一般的に、

独立変数（説明変数）が

質的変数のとき

⇒数量化Ⅱ類

量的変数のとき

⇒判別分析

とされる。

その他の分析との比較

	独立変数 （説明変数）	従属変数 （外的基準）
数量化Ⅰ類	質的変数	量的変数
数量化Ⅱ類	質的変数	質的変数
判別分析	量的変数	質的変数
重回帰分析	量的変数	量的変数

※どの分析も独立変数によって、従属変数を説明・分類・予測したりするもの。つまり、目的はほぼ同じと言え、補助的な解析処理を行えば、どれも同じ結果が得られることになる。

まとめ

数量化Ⅱ類および判別分析は、
過去の傾向や結果を利用し、
新しいデータがどのように分類されるかを
予測するために使われる。
つまり、何らかの判断基準をつくるときに有効といえる。

＜臨床場面への応用例＞

- 症状から病名を判断する
- 職業・進路の適正を判断する

参考

大村平　1985　多変量解析のはなし　日科技連

林知己夫著作集編集委員会・編　2004 林知己夫著作集第3巻　質を測る—数量化理論—　勉誠出版

森敏昭・吉田寿夫　1990　心理学のためのデータ解析テクニカルブック　北大路書房

渡部洋　1988　心理・教育のための多変量解析法入門〔基礎編〕　福村出版

URL：Researcher情報交流街

URL：おしゃべりな部屋（統計学：青木研究室）