URL: http://www.nik.sal.tohoku.ac.jp/~tsigeto/statb/
作成：田中重人（講師）<tsigeto@nik.sal.tohoku.ac.jp>

比較現代日本論研究演習 II

「多変量解析の基礎」
（2001 年度第 2 学期） 大学院生対象
＜水２＞コンピュータ実習室（文学部本館 7F 711-2）

授業の概要（予定）10/9 現在

目次

1. イントロダクション (10/10)
2. SPSS 入門、データ配布 (10/17)
3. 1 変量の分布 (10/24)
4. 2 変量の関連 (10/31)
5. 因子分析 (11/7～11/21)
6. 因子分析に関する論文講読 (11/28)
7. 回帰分析 (12/5～12/19)
8. 因子分析と回帰分析の応用 (1/9)
9. 回帰分析に関する論文講読 (1/16)
10. レポート作成

※（）内の日付はおおよその計画をあらわしているが、実際の授業の進行状況によって前後にずれることがある。

成績評価について

- ほぼ毎回課題を出します。たいていは授業中に作業をしてその場で提出してもらう形式にするつもりです。
- 学期末のレポートは、授業で配布するデータを使って、因子分析・回帰分析の両方を使った分析結果を自由に書いてください。(詳細はあらためて指示します)

教科書

- 大野高裕、1998『多変量解析入門』同友館、ISBN 4-496-02752-6。

生協 (文系書籍部) に入荷しているので、各自購入のこと。

その他の参考文献

- 宮脇 典彦＋和田 悟＋阪井 和男、2000『SPSS によるデータ解析の基礎』培風館、ISBN 4-563-00888-5。
- 古谷野 亘、1988『数学が苦手な人のための多変量解析ガイド』川島書店、ISBN 4-7610-0391-X。
- 三土 修平、2001『数学の要らない因子分析入門』日本評論社、ISBN 4-535-55217-7。

2001.10.10

比較現代日本論研究演習 II (田中重人)
受講登録フォーム

氏名：

学年：

学生番号：

所属（文学部日本語教育以外の場合）:

興味のあること（非学術的な話題も可）:

・自宅でパソコンを使えますか? **ある / ない**

・SPSS を使った経験がありますか? **ある / ない**

・コンピュータ・プログラムを作成したり、プログラミングの授業を受けたりしたことがありますか? **ある / ない**

ある場合 → 言語名（　　）

比較現代日本論研究演習 II
「多変量解析の基礎」

カードをとって
適当なところに着席

電源はまだ入れない

2001.10.10

比較現代日本論研究演習 II

多変量解析の基礎

東北大学大学院文学研究科 2001 年度
田中 重人（講師）

【目的】
因子分析・回帰分析の習得

- 多数の変数に共通する要因
  ＝因子分析
- 因果関係のモデル化
  ＝回帰分析

【教科書】

大野高裕、1998『多変量解析入門』同友館。

生協（文系書籍部）に入荷済

【授業の形式】

★ 講義＋実習
★ 論文講読(2 回)
★ 期末レポート

(授業予定は次のページ)

【授業の予定】

イントロダクション (10/10)
SPSS 入門、データ配布 (10/17)
1 変量の分布 (10/24)
2 変量の関連 (10/31)
因子分析 (11/7～11/28)
回帰分析と応用 (12/5～1/19)

【実習室について】

★ 入室には学生証が必要。
★ 土足・飲食・喫煙厳禁。
★ 退出時に必要事項を紙に記入。
（書けるところを書いてみよう）

★ コンピュータの起動と終了
（ディスプレイの電源を落とすのを忘れないこと）

★ ファイルの保存場所について

授業でつかうファイルは、授業開始時に My Document フォルダにコピーして使う。授業終了時に削除する。

内蔵 Disk に個人データをおいてはいけない

★ フロッピーを各自購入

「DOS フォーマット」(3.5 インチ)のものが便利。

【この授業で使用するデータ】
1995 年 SSM 調査の一部

cf. 『日本の階層システム』(全 6 巻)
東京大学出版会、2000 年。

【SPSS の起動】

★ 「スタート」→「プログラム」
→「SPSS for Windows 10.0J」
★ 「データに入力」をチェック
★ 「データエディタ」が開くことを確認

【変数の定義】

★ 「変数ビュー」タブに切り替え
★ 変数名を必要なだけつくる
★ 「データ ビュー」タブに切り替え
★ データを入力

適当な名前で保存していったん終了。
→ ○○○○.sav というファイルができる

1. データの配布
2. 標本抽出について
3. SPSS 入門
4. データの変換

【データの配布】

1995 年 SSM 調査 B 票の一部

★ 全国から 70 歳以下の有権者を層化 2 段無作為抽出

★ 訪問面接法

cf. 『日本の階層システム』（全 6 巻）東京大学出版会、2000 年。

★ 意識項目と基本的属性に限定

★ 250 ケースをランダムに抽出

★ 未公開のデータなので流出しないように

★ 変数ラベルは菅野剛(大阪大学) 氏による

【標本抽出の 4 段階モデル】

ユニバース (universe)

母集団 (population)

計画標本 (designed sample)

有効標本 (valid sample / case)

★ 伝統的な統計学では 4 段階にわけずに、2 段階で考えるのがふつう：

母集団=Universe + population

標本 = (designed/valid) sample

【無作為抽出】

母集団から計画標本を選ぶ際に、母集団にふくまれるすべての個体の抽出確率が等しくなるように抽出する (random sampling)

➡「等確率標本」

つぎの条件が必要：

★ 母集団の人口が既知

★ 個体を網羅した「台帳」

※ 個体によって抽出確率が違う場合も、事後的に調整して等確率標本と同様の統計処理をおこなうことは可能

※ 「台帳」が完備してない状況でも、工夫次第で無作為抽出に近づけることができる

【無作為抽出の実際】

★ 2 段抽出 ＝ 2 段階の抽出単位を設定

例：市町村→住民、学校→生徒

- 確率比例抽出法：その抽出単位が含む個体数に抽出確率を比例させる。
- 等確率抽出法：上位抽出単位の抽出確率は一定にしておき、個体の抽出数のほうを調整。

★ 系統抽出 ＝「台帳」から等間隔に抽出。

- スタート番号は乱数で決める
- 抽出間隔は次のことを考えてきめる
  (1) 台帳のもつ周期性と同調しない
  (2) 台帳全体をカバーできる
  具体的には 台帳人数／計画標本数に近い素数をえらぶのがよい。

★ **層化抽出法**＝母集団を層別にわけ、各層の人数に比例して標本数を割り当てる

・結果に影響を与えそうな重要な属性についておこなう：性別・年齢・地域など

・抽出単位や個体がどの層に属しているかを台帳から判断できないと使えない

※「層別抽出法」「比例割当抽出法」ともいう

実際の調査で理想的な標本抽出ができることはまずない。

また計画標本のなかから無効回答がでるので、

無作為ではない誤差がかならず発生する。

この誤差は**統計的には処理できない**ので、個別に推測する

・どの層を過剰に代表しているかを把握する
・おなじ母集団を対象にした調査と比較する

【宿題】

論文や新聞･雑誌記事で使われている調査データについて、標本抽出の 4 段階にそって紹介する。

【データ・セット】

★ ケース × 変数

★ 変数は変数名で管理

★ 変数名以外に「ラベル」

★ 無回答などの欠損値 (.)

【SPSS のウインドウ構成】

● データ・エディタ

● シンタックス・エディタ

● 出力ビューア

【メニューとシンタックス】

★ 分析手法をえらぶ

★ 必要なオプションを指定

★ 「貼り付け」をクリック

★ シンタックスの必要部分を選択して実行（▶）

【変数値の再割り当て】

データエディタのメニューバーで

● 「変換」→「値の再割り当て」→「他の変数へ」

● 変換先変数の名前をつける

● 「今までの値と新しい値」

● 値の組を指定したら「続行」

● シンタックスを貼付けて実行

● 新変数の度数分布を確認

● 問題がなければデータセットを保存する

【出力ビューア】

★ 左側に目次、右側に出力内容

★ エラー表示もここに出る

【印刷】

★ 左側の目次で選択

★ 印刷前にプレビュで確認

2001.10.24 比較現代日本論研究演習 II

第 3 回「1 変量の分布」(田中重人)

1. データ尺度の種類
2. 度数分布表とヒストグラム
3. 代表値と散布度
4. 平均と標準偏差

【データ尺度の種類】

●名義尺度 (nominal scale)
(質的変数とも)
●順序尺度 (ordinal —)
●間隔尺度 (interval —)
●比率尺度 (ratio —)

(教科書 p. 41–48)

★ 上位の尺度のほうが使える演算が豊富
★ 上位の尺度は下位の尺度の特徴を兼ね備えている

➡ 分析手法の選択幅が広い

私たちが測定するものは、
たいてい順序尺度以下である。

★上位の尺度への変換には
一定の理論的根拠が必要

【度数分布表】

Frequencies コマンドを使う

★ 度数
★ 相対度数 (%)
★ 累積度数・累積相対度数
★ 欠損値のあつかい

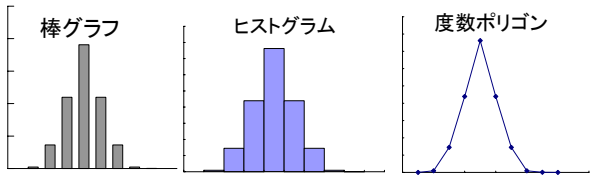
【棒グラフとヒストグラム】

●棒グラフ……棒同士の間に空白をあける。高さ (長さ) をよむ。
●histogram (柱グラフ)……柱の間隔をあけない。面積をよむ。

※縦軸は度数または%

★ 連続量を階級分けした場合
→ ヒストグラム
★ それ以外の場合 (質的変数／離散量) → 棒グラフ

※度数多角形 (polygon) は複数の変数の分布を比較するときに便利。



SPSS では histogram は書きにくい
★ recode で整形した上で度数分布表のメニューで「図表…」指定。棒グラフを書く
★グラフ→インタラクティブ→ヒストグラム では等間隔の区間に分割してくれる

【代表値と散布度】

★ 平均値 (mean) — 標準偏差 (SD)
(間隔尺度以上)

★ 中央値 (median) — 四分位偏差 (Q)
(順序尺度以上)

【平均値】

総和をデータ数で割ったもの

【標準偏差】

平均値からの偏差の２乗値の平均が「分散」
分散の平方根が「標準偏差」

★ 平均値と標準偏差はセットで使う

【SPSS のコマンド】

「記述統計」→「記述統計」

→ 変数とオプションを指定

【平均値を使うときの注意事項】

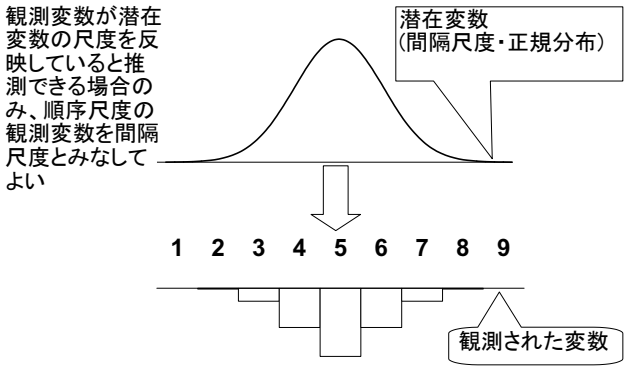
★平均値ははずれ値の影響を受けやすい。
あまりにかけはなれたケースがあるときは

- **上下数%を取りのぞいたデータセットで計算する (調整平均)**
- **順位に変換したり中央値を使って分析**

★平均値・標準偏差は間隔尺度以上のデータに対してしか意味をもたない。
順序尺度の平均値をとっていいのは

- **潜在的には間隔尺度のはず**
- **測定のポイントが一定間隔**

という２条件をともに満たす場合



➡ **具体的には**

- **4 点以上での測定している**
- **正規分布に近似している (教科書 p.15) :**
  - **単峰性**
  - **左右対称性 (歪度)**
  - **中央への集中度 (尖度)**

ヒストグラムを描いて検討するとよい。

正規分布との乖離度を統計的に検討する手法もある

➡ **これらの条件を満たさない場合は**

- **非線形変換 (対数・平方根など)**
- **順位に変換したり中央値を使って分析**

※ 間隔尺度のデータでも、
左右対称でないものについては
平均値よりも中央値のほうが
適当であることが多い

典型例：収入・人口など

【課題】

(1) 本人年齢の度数分布表を出力し、中央値と上側 25%点に印をつける

(2) 適当な変数について棒グラフまたはヒストグラムを作成し、横軸上に平均値と標準偏差を書き入れる

2001.11.7 比較現代日本論研究演習 II

第 5 回「2 変量の関連」(田中重人)

1. 変数の標準化
2. 相関係数
3. 相関係数行列
4. 欠損値の処理

【変数の標準化】

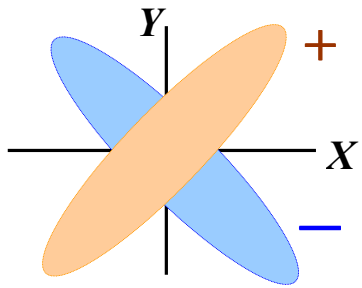
平均＝０，標準偏差＝1
になるように変換する。

$$X = \frac{x - \text{平均}}{\text{SD}}$$

これで単位を気にせずに比較できるようになる

【変数間の関係】

正(＋)の関係か、負(－)の関係か



【相関係数】

Pearson の積率相関係数 (教科書 p. 19)

(Product-moment correlation coefficient)

標準化済みの変数 X, Y について

$$r = \frac{XY\text{の総和}}{N}$$

今日では、単に「相関係数」といえばこの r をさす

r は －1～＋1 の範囲の値をとる:

- 無関連のときゼロ
- 完全に直線的な関連のとき±1

解釈の目安：
- 0.1 以下なら無関連、
- 0.2 程度で弱い相関、
- 0.4 程度で強い相関、
- 0.7 をこえると「ほとんどおなじ変数」

【SPSS コマンド】

「相関」→「2 変量」

変数を指定する

【相関係数行列】

3 変数以上の相関係数を総当たりで出すこともできる(correlation matrix)

ただし、「オプション」で「リストごとに除去」をえらぶ

【欠損値の処理】

- 対単位 (pairwise) の除去
  個々の組合わせごとに欠損ケースを除く
- 表単位 (listwise) の除去
  分析に使う変数に**ひとつでも**欠損のあるケースを除く

多変量解析の際は listwise 処理がふつう。

ただし、欠損ケースが多くなる。
1 割以上が欠損になるようなら注意。

【注意事項】

★ 相関係数は、はずれ値や歪みに弱い(平均値と同様)。順位相関係数 (Spearman, Kendall など) のほうがいいことも。

★ クロス表や散布図で**関連のかたち**をチェックしておくことがのぞましい

★ 相関係数の 95%信頼区間は
N=100 で±0.20, N=200 で±0.14 くらい

比較現代日本論研究演習II (田中 重人)

2001.11.7 課題

氏名:
学年:
所属:
学生番号:

(1) 次の変数を標準化せよ

	x	y
平均	2.61	2.08
SD	0.86	0.84
	1→	1→
	2→	2→
	3→	3→
	4→	4→

(2) つぎのクロス表をもとに相関係数を計算せよ

x	y			
	1	2	3	4
1	17	3	5	0
2	21	51	6	0
3	20	43	37	2
4	7	4	15	8

(N=239)

比較現代日本論研究演習II (田中 重人)

2001.11.7 課題

氏名:
学年:
所属:
学生番号:

(1) 次の変数を標準化せよ

	x	y
平均	2.61	2.08
SD	0.86	0.84
1→	**-1.87**	**-1.29**
2→	**-0.71**	**-0.10**
3→	**0.45**	**1.10**
4→	**1.62**	**2.29**

(2) つぎのクロス表をもとに相関係数を計算せよ

x	y			
	1	2	3	4
1	17	3	5	0
2	21	51	6	0
3	20	43	37	2
4	7	4	15	8

(N=239)

XY をもとめる:

X	Y			
	−1.29	−0.10	1.10	2.29
−1.87	**2.41**	**0.18**	**-2.05**	**-4.28**
−0.71	**0.91**	**0.07**	**-0.78**	**-1.62**
0.45	**-0.58**	**-0.04**	**0.50**	**1.04**
1.62	**-2.08**	**-0.15**	**1.77**	**3.69**

XYに各セルの人数をかける:

X	Y			
	−1.29	−0.10	1.10	2.29
−1.87	**40.92**	**0.53**	**-10.25**	**0.00**
−0.71	**19.15**	**3.45**	**-4.66**	**0.00**
0.45	**-11.66**	**-1.86**	**18.38**	**2.07**
1.62	**-14.55**	**-0.62**	**26.55**	**29.55**

総和 ＝	**97.01**
総和／N＝	**0.41**

2001.11.14 比較現代日本論研究演習 II

第 6 回「因子分析(1)」(田中重人)

1. 多変量解析のツボ
2. 類似関係型の分析
3. 因子分析の基礎
4. SPSS のコマンド

【多変量解析のツボ】

★ **目的** (類似関係か因果関係か?)
★ モデル構造 ($z=a_1x_1+a_2x_2+\cdots+a_nx_n$)
★ 係数のポリシー (z の分散を最大化)
★ 係数の算出 (相関行列の固有ベクトル)
★ **結果の検討** (Fit and meaning)

(教科書 p. 56–61)

【ふたつの目的】

● **類似関係型**
因子分析，クラスター分析……

● **因果関係型**
回帰分析，判別分析……

(教科書 p.48–56)

【類似関係型の分析】

因子分析 (factor analysis)
……間隔尺度の相関行列を使う
- 主成分法(principal component)
  ……特別にあつかって「主成分分析」とも
- 主因子法(principal factor)
- 最尤法 (maximum likelihood)
- その他

クラスター分析 (cluster analysis)
……さまざまな尺度の(非)類似行列を使う

類似関係型の分析でできること：

★ 似た変数同士をまとめる
➡ cluster

★ 潜在的要因を抽出
➡ factor

★ 少数の変数に縮約
➡ component, axis, vector, score…

【因子分析の基礎】

(この授業では主成分法に限定する)

2 変数の分布を「うまく説明する」直線

各点からの距離がいちばん小さい
‖
その直線上に投影したときの SD が最大

この直線を**「主成分」**という

【固有値】

固有値 (eigenvalue)
＝主成分の分散 (SD^2)

★ 固有値の最大値は変数の個数

★ 2 変数の場合は、固有値 ＝ 1 + $|r|$

【寄与率】

寄与率 ＝固有値／変数の個数

★ 寄与率は 0～1 の範囲の値をとる

【SPSS コマンド】

「データの分解」→「因子分析」

★ 変数を指定する。
★「記述統計」で「相関行列」の「係数」をチェック
★「オプション」で欠損値が「リストごとに除外」になっていることを確認

2001.11.21 比較現代日本論研究演習 II

第 7 回「因子分析(2)」(田中重人)

1. 因子分析の考えかた
2. 因子数の決定
3. Varimax 回転
4. 因子負荷量
5. 変数の共通性

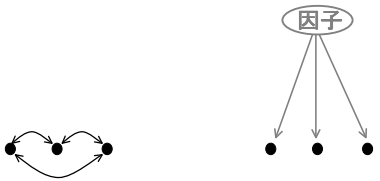
【因子分析の考えかた】

直接観測できない潜在的な因子

↓

表面的には測定値間の相関としてあらわれる

相関行列 ➡ 因子分析



【主成分法の因子分析】

主成分＝因子 と考えるのが「主成分法」

※ 因子を求める方法はほかにもいろいろある

【因子数の決定】

主成分は変数の数だけ抽出できる

(第 1 主成分……第 n 主成分)

★ あとになるほど固有値 (寄与率) が小さくなる

★ 全主成分の寄与率の合計 (累積寄与率) は 1 になる

それらの一部だけを使う = 因子数の決定

★因子数が少ないほど単純になる

★因子数が多いほどデータへの あてはまりがよくなる

【因子数決定の基準】

固有値が 1 以上の主成分のうち

★「肘」の手前まで (scree plot)

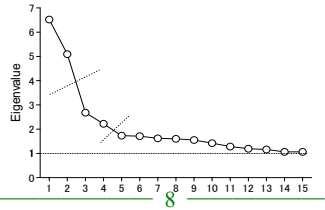
★累積寄与率が一定の値 (例えば 0.5) まで

複数試してみて、解釈が容易でデータへの当てはまりがよいものをえらぶ

【スクリープロット】

Scree plot

固有値または寄与率の折れ線グラフ



【SPSS コマンド】

「データの分解」→「因子分析」

★ 変数、「相関行列」の「係数」を指定

★ 「因子抽出」→「スクリープロット」

出力を見て、因子数を決めてやりなおす

★「因子抽出」の「抽出の基準」で「因子数」を直接指定

【Varimax 回転】

主成分 (初期因子) は全変数を均等に代表する傾向がある

そのままでは解釈しにくいので、変数のまとまりを反映するように回転 (rotate)させる

(→単純構造)

Varimax はもっともポピュラーな回転法

(ほかにもいろいろある)

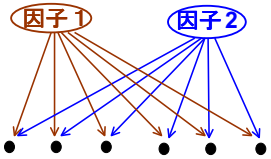
★ SPSS では、「回転」で「方法」＝「バリマックス」を指定

★ ついでに変数のならべかたを指定：「オプション」の「係数の表示書式」を「サイズによる並び替え」に

【因子負荷量】

factor loadings / factor pattern

因子と変数との相関係数 (のようなもの)



※ SPSS では「成分行列」と呼ばれている

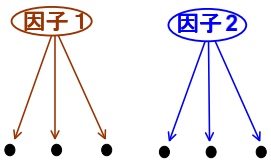
【単純構造】

因子分析に投入したすべての変数について、

★ある因子についてだけ負荷量が大きく

★ほかの因子については負荷量はほぼゼロ

であるとき、この因子負荷量行列は「単純構造」(simple structure) である



➡ 対応する変数だけを使ってそれぞれの因子を抽出してもほとんどおなじ結果になる

(本当はすこしちがう)

【変数の共通性】

communality

抽出した因子によって各変数がどれくらい説明されているか

= その変数についての負荷量の 2 乗和

共通性の低い(例えば 0.3 未満) 変数は除いてしまったほうがよいことが多い

【指数表記の読みかた】

SPSS の出力で 1.346E-02 などとあるのは、スペースを節約して小さい数値を書く方法：

$1.346\ \mathrm{E}\ {-02} \ = \ 1.346 \ \times \ 10^{-2} \ = \ 0.01346$

「E の後ろの数だけ小数点をずらす」と考えるとよい

因子分析の出力例（部分）

共通性

		初期	因子抽出後
Q27A	評価高い職業	1	0.54
Q27B	高い収入	1	0.46
Q27C	高い学歴	1	0.54
Q27D	家族の信頼尊敬	1	0.41
Q27E	volunteer、町内会活動	1	0.42
Q27F	趣味サークル	1	0.27
Q27G	多くの財産	1	0.56
Q27H	高い地位	1	0.65
Q28A	他人に追い越されそう	1	0.16
Q28B	獲得したものを維持	1	0.20
Q28C	がんばっている	1	0.47
Q28D	心の豊かさ	1	0.48
Q28E	自分には良い点があり	1	0.15

因子抽出法: 主成分分析

説明された分散の合計

	初期の固有値			抽出後の負荷量平方和			回転後の負荷量平方和		
成分	合計	分散の%	累積%	合計	分散の	累積 %	合計	分散の	累積 %
1	3.52	27.05	27.05	3.52	27.05	27.05	3.11	23.96	23.96
2	1.82	13.97	41.02	1.82	13.97	41.02	2.22	17.06	41.02
3	1.26	9.67	50.68						
4	1.06	8.19	58.87						
5	0.90	6.95	65.82						
6	0.85	6.52	72.34						
7	0.71	5.50	77.84						
8	0.63	4.86	82.70						
9	0.55	4.26	86.95						
10	0.50	3.82	90.78						
11	0.43	3.34	94.12						
12	0.41	3.14	97.26						
13	0.36	2.74	100.00						

因子抽出法: 主成分分析

回転後の成分行列

成分	1	2
高い地位	**0.81**	-0.02
多くの財産	**0.75**	0.06
評価高い職業	**0.73**	0.13
高い学歴	**0.72**	0.14
高い収入	**0.67**	0.10
他人に追い越されそう	**0.40**	0.07
趣味サークル	0.39	0.35
心の豊かさ	-0.08	**0.69**
がんばっている	0.06	**0.68**
家族の信頼尊敬	0.11	**0.63**
volunteer、町内会活動	0.25	**0.60**
獲得したものを維持	0.08	**0.44**
自分には良い点があり	0.05	0.38

因子抽出法: 主成分分析　　回転法: Kaiser の正規化を伴うﾊﾞﾘﾏｯｸｽ法

a　　3回の反復で回転が収束しました。

2001.11.28 比較現代日本論研究演習 II

第 8 回「因子分析(3)」(田中重人)

1. 単純加算得点
2. 因子得点
3. 因子得点の性質
4. 結果のプレゼンテーション

【単純加算得点】

単純加算得点 ＝ ある因子への負荷量の高い変数を集め、それらを加算した変数をつくる

SPSS では、データ・エディタの「変換」メニューから「計算」をえらび、

目標変数 ＝ ○○ ＋ ○○ ＋ ・・・

のように指定する

【因子得点】

factor score (教科書 p. 106–107, 142–143)

個々のケースについて、因子の値を計算

➡変数として保存して、ほかの分析に使うことができる

【SPSS コマンド】

★ 因子分析メニューの「得点」オプションで「変数として保存」をえらぶ。方法は「回帰法」にする

★ データ・エディタで確認

★ 「変数ビュー」でラベルか変数名を変更

★ 「ファイル」→「名前を付けて保存」

【因子得点の性質】

(主成分法、Varimax 回転の場合)

★ 平均=0, SD=1

★ 因子得点同士の相関はゼロ

★ 各変数との相関 ＝ 因子負荷量

因子得点は

(1)標準化した変数値に

(2) 負荷量をもとに算出した係数をかけて

(3)それらを足し合わせたもの

SPSS では因子分析の「得点」オプションで「因子得点係数行列を表示」をえらぶと、(2)の係数がわかる

【単純加算得点と因子得点】

	もとの変数	係数
単純加算	そのまま	1 か 0
因子得点	標準化	負荷量から計算

変数がおなじ尺度で測定されていて単純構造の因子負荷量の場合、単純加算得点と因子得点と 相関≒1 になる

【結果のプレゼンテーション】

因子分析の結果提示に必要な情報

(1) もとの変数の記述統計量
(平均・SD・欠損ケース数)

(2) 全成分の固有値 (寄与率) の一覧

(3) 回転後の因子負荷量行列・共通性・固有値・寄与率

相関係数行列を提示するのもよい

ただし

(1) は通常リストワイズ処理にしない。
共通性が低いなどのために除いた変数についても表示するのがよい。
また、尖度や歪度を表示してもよい。

(2) は、変数が多いときは、固有値 1 以上の成分に限ってもよい

比較現代日本論研究演習 II (田中重人)
第 8 回「因子分析(3)」(2001.11.28) 資料

表 1 分析に使う変数の記述統計量

	平均	SD	歪度	尖度	有効数	欠損数
問 27 次にあげることがらは、あなたにとってどのくらい重要ですか						
Q27a 社会的評価の高い職業につくこと	2.614	0.859	-0.160	-0.589	241	9
Q27b 高い収入を得ること	2.073	0.841	0.318	-0.631	245	5
Q27c 高い学歴を得ること	2.661	0.880	-0.191	-0.647	245	5
Q27d 家族から信頼と尊敬を得ること	1.404	0.624	1.389	1.269	245	5
Q27e ボランティア活動、町内会活動など社会活動で力を発揮すること	2.034	0.861	0.535	-0.331	238	12
Q27f 趣味やレジャーなどのサークルで中心的役割を担うこと	2.653	0.832	-0.190	-0.485	242	8
Q27g 多くの財産を所有すること	2.597	0.858	-0.129	-0.599	248	2
Q27h 高い地位につくこと	2.893	0.813	-0.502	-0.081	242	8
問 28 あなたにとって次のような気持ちや考えはどの程度あてはまりますか						
Q28a まごまごしていると、他人に追い越されそうな不安を感じる	3.438	1.194	-0.353	-0.878	249	1
Q28b もっと多くを手にするよりも、これまでに獲得したものを維持することの方が重要であると思う	2.657	1.137	0.197	-0.806	245	5
Q28c 日頃の生活で、私は自分なりによくがんばっていると思う	2.145	1.037	0.690	-0.135	249	1
Q28d これかれは、物質的な豊かさよりも心の豊かさやゆとりのある生活をすることに重きをおきたいと思う	1.738	0.820	0.963	0.580	248	2
Q28e 自分には多くのよい点があると思う	2.727	0.888	0.062	0.009	242	8

選択肢はつぎのとおり：問 27 は「1. 重要である」「2. やや重要である」「3. あまり重要ではない」「4. 重要ではない.」; 問 28 は「1. よくあてはまる」「2. ややあてはまる」「3. どちらともいえない」「4. まったくあてはまらない」「5. まったくあてはまらない」。

表 2 因子分析の結果

(a) 初期解の固有値と寄与率（主成分法）

成分	固有値	寄与率 (%)	累積寄与率 (%)
1	3.097	34.4	34.4
2	1.675	18.6	53.0
3	0.911	10.1	63.1
4	0.716	8.0	71.1
5	0.636	7.1	78.2
6	0.591	6.6	84.7
7	0.559	6.2	90.9
8	0.454	5.0	96.0
9	0.362	4.0	100.0

使用した変数は (b) を参照。

(b) 採択した解（主成分法、Varimax 回転）

	因子 1	因子 2	共通性
Q27h 高い地位	0.806	-0.025	0.651
Q27c 高い学歴	0.742	0.167	0.578
Q27g 多くの財産	0.741	-0.005	0.549
Q27a 評価高い職業	0.741	0.193	0.586
Q27b 高い収入	0.696	0.116	0.498
Q28d 心の豊かさ	-0.085	0.714	0.518
Q27d 家族の信頼尊敬	0.123	0.676	0.473
Q27e volunteer や町内会活動	0.204	0.661	0.478
Q28c 私はがんばっている	0.084	0.659	0.442
固有値	2.854	1.918	4.772
寄与率 (%)	31.7	21.3	53.0

質問文は表 1 を参照。

2001.12.12 比較現代日本論研究演習 II

第 10 回「重回帰分析(1)」(田中重人)

1. 因果関係の設定
2. 回帰分析とは
3. 最小 2 乗法
4. SPSS コマンド
5. 標準化係数

【因果関係の設定】

目的変数 (dependent variable)

結果になる変数 (ひとつ) : 従属変数とも

説明変数 (independent variable)

原因になる変数 (複数可) : 独立変数とも

目的変数と説明変数はしばしば Y と X であらわされる

【因果関係の設定のルール】

★ 時間的な順序関係
★ (実験の場合) 操作の順序
★ 先行研究でのあつかい
★ 一般的常識

つまるところ、絶対的なルールはない
→分析者が恣意的に決めるもの

【回帰分析とは】

Regression analysis

X の値によって Y が決まる、と考えて説明する

- Y をうまく説明できるような「回帰直線」を引く (最小 2 乗法)
- 直線のパラメタ (とくに傾き) を評価する (回帰係数)
- 回帰直線からのずれを評価する (決定係数)

【最小 2 乗法】

ordinal least square method

適当な直線 A + BX によって Y の値を近似する方法。

Y と A+BX とのずれの大きさを評価するために
差の 2 乗和をとる。

この 2 乗和 $\Sigma(Y-A-BX)^2$ が最小になるように

A と B の組み合わせを求める。

※ X と Y を入れ替えると結果が変わることに注意

A を「**定数**」または「**切片**」(intercept),
B を「**回帰係数**」(regression coefficient) という。

回帰係数 B の意味 :
X が 1 単位増えたとき Y がどれだけ増えるか

【SPSS コマンド】

「分析」→「回帰」→「線型」

★ 従属変数と独立変数を指定する

→ 「係数」の出力を見る
「非標準化係数」が B
「定数」が A にあたる

【標準化係数】

X と Y の両方を標準化したうえで回帰分析をおこなった場合の係数を「標準化係数」(standardized coefficient) といい、ギリシャ文字の β であらわす。

β の値は、X と Y との相関係数に等しい

SPSS では標準化係数はデフォルトで出力される

2001.12.19 比較現代日本論研究演習 II

第 11 回「重回帰分析(2)」(田中重人)

1. 重相関係数・決定係数・寄与率
2. 重回帰分析
3. 標準偏回帰係数の比較
4. 標本誤差の推測

【重相関係数】

multiple correlation coefficient

実績値と回帰式による予測値との相関。

大文字の R であらわす。

0~1 の範囲の値をとる。

(教科書 p. 75)

R^2 を「決定係数」という。(＝寄与率)

【寄与率】

ふたつのモデルを比較する：

・平均値モデル：Y = M

・回帰式モデル：Y = A + BX

それぞれの残差の平方和 (2 乗和) を計算

(SS: Sum of Squares)

ちなみに平均値モデルでは $SS = SD^2 \times N$

$$\text{寄与率} = 1 - \frac{\text{回帰式モデルSS}}{\text{平均値モデルSS}} = R^2$$

寄与率は 0~1 の範囲の値をとり、

Y のばらつきのうち何%が

X によって説明できるかを表す。

SPSS では「モデル集計」の項に R と R^2,
「分散分析」の項に SS (全体と残差)
が表示される

【重回帰分析】

multiple regression analysis

複数の説明変数を投入した回帰分析

$$Y = A + B_1X_1 + B_2X_2 + \cdots$$

係数の求めかたは単回帰分析とおなじ (OLS)

【偏回帰係数】

各変数にかかる係数 B_i の意味：

「他の変数を一定の値に固定したとき、その変数の１単位の増加が Y をどれだけ増やすか」

特に「偏回帰係数」(partial~) と呼ぶことがある。

(cf. 偏微分 partial differentiation)

※「偏相関係数」とはちがうので注意

もし説明変数間に相関がなければ、単回帰分析の係数と重回帰分析の係数は等しい。

説明変数間に相関がある場合は、重回帰分析のほうが係数がちいさくなることが多い

(例外もあり)

相関が大きい変数を投入するのはよくない

(多重共線性：multi-co-linearity)

およそ $r < 0.7$ が条件

【標準偏回帰係数】

単位のちがう説明変数を投入した場合、

そのままでは係数を比較しにくい

↓

標準偏回帰係数 (β) は比較可能 (教科書 p. 79)

各変数の Y への影響力 (effect) をあらわすと解釈できる

Q: 相関係数との関係は?

【標本誤差の推測】

調査でえられたデータは、母集団の一部の標本についてのものである。

もう一度標本をとりなおして分析したら

ちがう結果が出るかも

→標本抽出にともなう誤差 (sampling error) を評価しておく必要がある

無作為抽出の場合の標本誤差

★　方向性をもたない

★　確率的に決まる

★　**標本数が大きいほど誤差の範囲が小さい**

➡「統計的推測」によって範囲を推定できる

【標準誤差】

Standard Error (SE)

偏回帰係数の標本誤差の大きさをあらわす。

3 つの要因できまる：

- **X, Y それぞれの標準偏差**　(大きいほど大きくなる)
- **分析に投入した標本数**　(大きいほど小さくなる)
- **説明変数同士の相関**　(大きいほど大きくなる)

【t 値】

B / SE を t 値という。

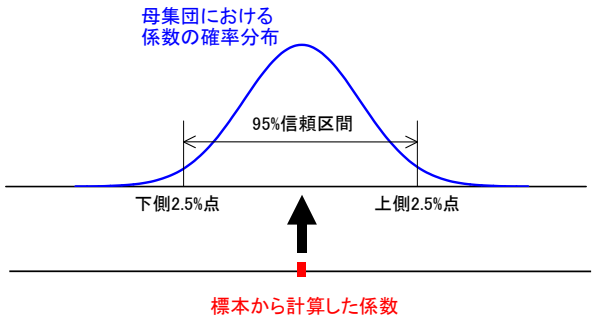
t 値は t 分布という確率分布にしたがうことがわかっているので、これをつかって、母集団における B の分布を推定できる。

【信頼区間】

母集団における係数の確率分布から両端を α%分だけ切り落としてえられる区間を $(100-\alpha)$%の「**信頼区間**」(confidence interval) という。

α を「**危険率**」、$(100-\alpha)$ を「**信頼率**」という。
この値は自由に決めていいのだが、通常は
$\alpha=5\%$ として、95%信頼区間を求める。

信頼区間のもとめかた



★　95%信頼区間のおおよその値：

$$B \pm 2.0 \times SE$$

B：係数　　SE：標準誤差　　2.0：t 臨界値

SPSS の重回帰分析では「統計」で
「係数の信頼区間」をチェック

【t 検定】

すべての係数について信頼区間を表示するのはわずらわしいので、通常は「信頼区間内にゼロがふくまれているか」だけを問題にする

たとえば 95%信頼区間に

- ゼロをふくむ → **5%水準で非有意**
- ふくまない　→ **5%水準で有意**

【有意確率】

significance level

危険率をどこまでさげると
ゼロがふくまれるようになるか

例：有意確率が 0.021 だとすると、

97.9%信頼区間の端がちょうどゼロにあたることになる

→5%水準では有意、1%水準では非有意

【F 検定】

決定係数 R^2 についても同様の統計的推測をおこなうことができる。→F 検定

ただし B の場合のような左右対称の信頼区間にならないので、計算がややこしい。

2002.1.16 比較現代日本論研究演習 II

第 12 回「重回帰分析(3)」(田中重人)

1. 表に書くべき事項
2. 余裕があれば書くべき事項
3. 試験とレポートについて

【表に書くべき事項】

※パラメタの推定値（回帰係数および定数）

※標準誤差

● t 検定の結果（推定値に * などをつける）

●標準回帰係数

●決定係数と F 検定の結果（* など）

●ケース数

※ 説明変数の単位に特に意味がなければ省略してよい

【表の例】

表 1 階層帰属意識の重回帰分析

説明変数	係数	(標準誤差)	β
(定数)	6.787**	(0.570)	ó
性別	0.032	(0.203)	0.011
満年齢	0.002	(0.007)	0.015
家族収入	−0.117**	(0.029)	−0.279

R^2=0.079**。N=238。

目的変数：階層帰属の主観的評価（10 段階，逆転）

**：1%水準で有意。*: 5%水準で有意。

無印：5%水準で非有意。β：標準偏回帰係数。

★ 小数第 2 位か第 3 位まで

★ 小数点をそろえる

【余裕があれば書くべき事項】

(1) 説明変数・目的変数の記述統計量
(平均・SD・欠損ケース数・尖度・歪度)

(2) 相関係数行列

【試験】

来週 (1/23) 試験をおこないます。
すでに配布済みの論文について、内容を理解できているかどうかを問うものです。

★何でも持ち込み可。

★各自、論文をよく読んでおくこと。

★田中のところに事前に論文の内容について質問にくることは禁じます。

【レポート】

締切：2/8 (金曜) 12:00

提出先：田中のレターケース、
または直接手渡し

課題：因子分析・回帰分析をおこなって、その結果を表にまとめ、コメントを書く。

★一連の分析にしても、別々の分析でもかまいません

★必要な事項が表に盛り込めているか、適切に解釈してコメントをかけているかを基準に評価します

★事前の相談可

比較現代日本論研究演習 II　(田中重人)

期末試験 (2002.1.23)

配布済みの論文「学校五日制に関する母親の意見の形成基盤」(轟亮 1995) を読み、下記の問題に答えよ。

【回答上の注意】

① 他の回答者の画面が見えないよう、互いに離れて座ること
② コンピュータで回答を書き、印刷して提出
③ 何を持ち込んで参照してもよいが、人に相談してはならない

1. この調査の設計について
   ① ユニバース (universe) は何か
   ② 標本 (sample) の抽出法を説明せよ

2. 表 2 では、学校五日制に関する意見について相関分析をおこなっている。この分析法をこのデータに適用することの問題点について述べよ。

3. 教育分業意識の因子分析について
   ① スクリープロット (scree plot) を作成せよ (Excel を使用)
   ② 単純構造 (simple structure) から外れた因子負荷量をもつ変数をひとつあげよ
   ③ 表 3-3 には必要な情報が欠けている。それは何か

4. 表 4 の相関係数行列では「父親職業威信」と「五日制への賛成」の関連は $r = 0.239$ だが、表 5 の重回帰分析では $\beta = 0.137$ と小さくなっている。こうなった原因を説明せよ。

(注) 「職業威信」とは、種々の職業に対する評価を多数の人に評定してもらい、その評定値の平均をつかって職業を序列づけたスコアである。日本では、直井 (1979) によって 1975 年に行われた調査によるスコアがよく使われる：

直井 優 (1979)「職業的地位尺度の構成」編= 富永 健一『日本の階層構造』東京大学出版会：p. 434–472。

具体的な値は、同書の巻末付表か http://www.nik.sal.tohoku.ac.jp/~tsigeto/ssm/occtable.txt を参照

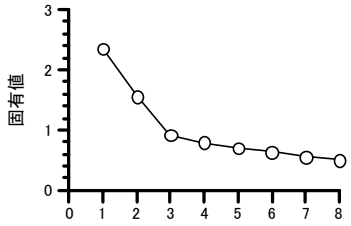
比較現代日本論研究演習 II　(田中重人) **期末試験** (2002.1.23)

解答例

1. この調査の設計について
   ① ユニバース (universe) ……**日本全国の公立学校に就学中の子どもとその親** (p.79)
   ② 標本 (sample) の抽出法……
   県立高校 4 校 (島根県東部の進学校，島根県西部の工業科高校，島根県山間部の普通科高校，石川県能登地方の普通科高校) から各 3, 6, 3, 9 クラスを抽出し、それらのクラスの全生徒とその保護者 (男女各 1 名) を標本とした。(p. 80)

2. 表 2 の問題点……**高校生の意見が偏っている (表 1-3) のと、父親の意見が単峰形でない (表 1-2) ので、そのまま間隔尺度を前提とした分析をするのはよくない**

3. 教育分業意識の因子分析について
   ① スクリープロット (scree plot)…… 表 3-2 から



縦軸は寄与率でもよい

   ② 単純構造 (simple structure) から外れた因子負荷量をもつ変数……**(f) 健康を保つ** または **(b) 人生について考える (表 3-3)**
   ③ 表 3-3 には必要な情報が欠けている。それは何か……**回転後の因子の寄与率**

4. 表 5 の β が小さくなっている原因……**互いに相関をもつ変数をコントロールしたため。おそらく「生活教育期待」と「母親学歴」の影響がおおきい。**