１．実験計画法【第１回】

1.1　１元配置実験計画法

多群間の平均や中間値に差があるかどうか検討する手法であり、いくつの変数を同時に比較するかによって１元配置（１元比較ともいう)、２元配置（２元比較ともいう）などと分かれている。ここでは理解し易い１元配置についてのみ解説する。

多群間の等分散の検定には Bartlett 検定を利用する。

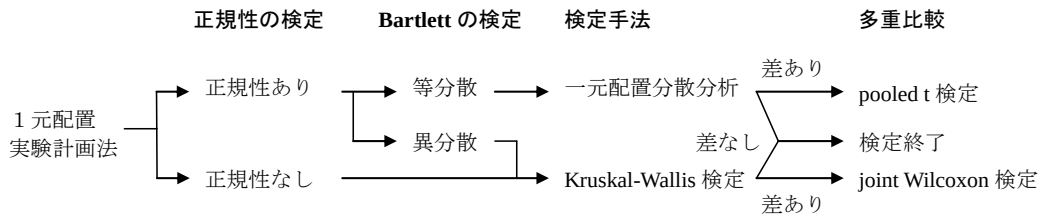


図 1.1　１元配置実験計画法の構造

例

３つの条件である商品の売上を調査したところ、実験計画法.txt (p1)の結果を得た。各群に差があるといえるか、実験計画法を用いて有意水準 5%で判定せよ。

正規性の検定　　正規分布と［みなす・いえない］
等分散性の検定　　検定確率［　　］　等分散と［みなす・いえない］
検定名［　　］　検定確率［　　］
判定　条件間に差があると［いえる・いえない］

差があるとするとどの条件間に差があるか。差がある条件同士を条件２＜条件３（これは実際の結果とは関係ない）のように不等号で表せ。

検定名［　　］
結果［　　］

問題１

実験計画法.txt (p2) は３つの工場群の不良品率を与えたものである。各群に差があるといえるか、実験計画法を用いて有意水準 5%で検討せよ。

正規性の検定　　正規分布と［みなす・いえない］
等分散性の検定　　検定確率［　　］　等分散と［みなす・いえない］
検定名［　　］　検定確率［　　］
判定　工場群間の不良品率に差があると［いえる・いえない］

差があるとするとどの条件間に差があるか。差がある条件同士を工場２＜工場３（これは実際の結果とは関係ない）のように不等号で表せ。

検定名［　　］

結果［　　］

問題２

実験計画法.txt (p3) は４つの群のデータであるが、各群に差があるといえるか、実験計画法を用いて有意水準 5%で検討せよ。

正規性の検定　　正規分布と［みなす・いえない］

等分散性の検定　　検定確率［　　］　等分散と［みなす・いえない］

検定名［　　］　検定確率［　　］

判定　群間に差があると［いえる・いえない］

差があるとするとどの群間に差があるか。差がある群同士を群２＜群３（これは実際の結果とは関係ない）のように不等号で表せ。

検定名［　　］

結果［　　］

問題３

実験計画法.txt (p4) は３群のデータであるが、各群に差があるといえるか、実験計画法を用いて有意水準 5%で検討せよ。

正規性の検定　　正規分布と［みなす・いえない］

等分散性の検定　　検定確率［　　］　等分散と［みなす・いえない］

検定名［　　］　検定確率［　　］

判定　群間に差があると［いえる・いえない］

差があるとするとどの群間に差があるか。差がある群同士を群２＜群３（これは実際の結果とは関係ない）のように不等号で表せ。

検定名［　　］

結果［　　］

演習【第２回】

ある４つの中学について英語・数学・国語の試験結果を調べた。多変量演習 1.txt のデータを読み込んで、以下の質問に答えよ。但し、検定は有意水準 5%とすること。

１．中学　　1）A 中学　2）B 中学　3）C 中学　4）D 中学

２．英語点数

３．数学点数

４．国語点数

1）数学について、各中学の平均（中央）値に差があるといえるか。

検定名［　　］　検定確率［　　］

判定　平均（中央）値に差があると［いえる・いえない］。

2）数学について各中学の平均（中央）値に差があるとすると、A, B, C, D どの中学の間に差があるか調べて A<B のように不等号で表せ。（差がある場合のみ）

検定名［　　］

結果［　　］

3）国語について、各中学間の平均（中央）値に差があるといえるか。

検定名［　　］　検定確率［　　］

判定　平均（中央）値に差があると［いえる・いえない］。

4）国語について、各中学間の平均（中央）値に差があるとすると、A, B, C, D どの中学の間に差があるか調べて A<B のように不等号で表せ。（差がある場合のみ）

検定名［　　］

結果［　　］

5）３教科の平均（中央）値に差があるといえるか。対応は考えないものとせよ。

検定名［　　］　検定確率［　　］

判定　平均（中央）値に差があると［いえる・いえない］。

6）３教科の平均（中央）値に差があるとすれば、どの教科の間に差があるか調べて英語＜数学のように不等号で表せ。（差がある場合のみ）

検定名［　　］　結果［　　］

1.2　対応がある場合の１元配置問題

例

３つの条件である商品の売上を調査したところ、実験計画法.txt（p1）の結果を得た。各データに対応があるとして差があるか検定せよ。

条件１　115，110，108，114，120，116，108，112，115，122
条件２　121，118，124，117，119，130，121，115，118，119
条件３　116，112，120，111，112，108，114，119，104，113

注）正規性の有無により、（繰り返しのない）２元配置分散分析か Friedman 検定を利用する。これは repeated measured 1 元配置分散分析、repeated measured Kruskal-Wallis 検定とも呼ばれている。

正規性の検定　　正規分布と［みなす・いえない］
検定名［　　］　検定確率［　　］
判定　群間に差があると［いえる・いえない］

1.3　２元配置分散分析

例

実験計画法.txt (p5) は作物の品種と肥料の組み合わせによる収穫量を表したデータである。2 元配置分散分析を用いて判定せよ。

注）2 元配置分散分析では、分類データの組み合わせによる交互作用が分かる。

品種水準間
　検定確率［　　］　水準間に差があると　［いえる・いえない］
肥料水準間
　検定確率［　　］　水準間に差があると　［いえる・いえない］
交互作用
　検定確率［　　］　交互作用に差があると［いえる・いえない］

２．重回帰分析【第３回】

例

重回帰分析.txt (p1)をもとに体重を身長と胸囲の１次関数で予測・説明する。

体重	身長	胸囲	体重	身長	胸囲
61.0	167.0	84.0	49.5	164.7	78.0
55.5	167.5	87.0	61.0	171.0	90.0
57.0	168.4	86.0	59.5	162.6	88.0
57.0	172.0	85.0	58.4	164.8	87.0
50.0	155.3	82.0	53.5	163.3	82.0
:	:	:	:	:	:
49.5	160.4	84.9	56.0	172.0	82.0

解説

体重 ＝ b_1 身長＋b_2 胸囲＋b_0　　の形で体重を予測する。

目的変数：体重　　説明変数：身長，胸囲

係数の値は？　　→　偏回帰係数

説明変数の重要性は？　　→　標準化偏回帰係数

どの程度予測できるか？　　→　重相関係数，寄与率（決定係数）

このモデルは有効か？　　→　Ｆ検定値と確率（要残差正規性）

それぞれの係数は有効か？　　→　ｔ検定値と確率（要残差正規性）

他の変数の影響を除いた目的変数と各説明変数の相関は？　　→　偏相関係数

どの程度予測できているのか図的に見たい　　→　散布図

どの程度予測できているのかデータ毎に見たい　　→　予測値と残差

問題１

重回帰分析.txt (p2)について、重回帰分析を行い、以下の問いに答えよ。

１）回帰式を求めよ。

卒業試験＝　［　　］入試点数　＋［　　］内申点数

＋［　　］勉強時間　＋［　　］出席率

＋［　　］

２）この回帰式の寄与率を求めよ。［　　］

３）この場合残差の分布は正規分布といえるか。［正規分布・正規分布でない］

変数増減法を用いて変数を自動選択する。

4）最終的な回帰式はどのようになるか。不要な変数の係数欄は空欄のままでよい。

卒業試験＝　［　　］入試点数　＋［　　］内申点数

＋［　　］勉強時間　＋［　　］出席率

＋［　　］

5）上の回帰式の寄与率を求めよ。［　　］

6）上の回帰式の寄与率はすべての変数を使った場合に比べ大きく下がっているか。

［大きく下がっている・あまり下がっていない］

7）この式を新しい予測モデルとして採用するか。　　［採用する・採用しない］

8）新しい予測モデルで、データ中の最初（1 番）の学生について卒業試験の実測値，その予測値，残差（実測値と予測値の差）はいくらか。

実測値［　　］　予測値［　　］　残差［　　］

9）上と同様のモデルで、質問項目の値が入試点数 70、内申点数 3.5、勉強時間 5、出席率 70%の学生の卒業試験はいくらに予測されるか。　　［　　］

演習 1

1）多変量演習 3.txt で、全変数を使った以下の重回帰式はどのように与えられるか。

試験成績＝［　　］×評定平均＋［　　］×模試１

［　　］×模試２＋［　　］×模試３＋［　　］

2）この重回帰式の寄与率はいくらか。［　　］

変数自動選択で偏回帰係数が有効である回帰モデルを作り、以下の問いに答えよ。

3）重回帰式はどのようになるか。説明変数に含まれないものは空欄のままにすること。

試験成績＝［　　］×評定平均＋［　　］×模試１

［　　］×模試２＋［　　］×模試３＋［　　］

4）寄与率はいくらになったか。［　　］

5）上の重回帰式を新しい予測モデルにして良いと思うか。［思う・思わない］

以後、新しいモデルで答えること。

6）データの中で最初の学生の予測試験成績はいくらか。［　　］

7）新しい重回帰式を利用すると以下の点数の学生の試験成績は何点に予測されるか。

変数名	評定平均	模試１	模試２	模試３
成績	3.5	70	73	75

予測試験成績［　　］

演習　重回帰分析【第４回】

目的変数＝b_1説明変数１＋b_2説明変数２＋・・・＋b_0　　の形で予測する。

係数の値は？　　→　偏回帰係数

説明変数の重要性は？　　→　標準化偏回帰係数

どの程度予測できるか？　　→　重相関係数，寄与率

このモデルは有効か？　　→　Ｆ検定値と確率（要残差正規性）

それぞれの係数は有効か？　　→　ｔ検定値と確率（要残差正規性）

どの程度予測できているのか図的に見たい　　→　散布図

どの程度予測できているのかデータ毎に見たい　　→　予測値と残差

演習２

多変量演習 4.txt のデータは各質問項目について 5 段階評価で、講義ごとに平均を取ったものである。重回帰分析を用いて以下の問いに答えよ。

総合評価を調査数以外のすべての変数で予測する重回帰モデル

１）回帰式を求めよ。

総合評価＝　［　　］進む速さ　＋［　　］声の大きさ

＋［　　］黒板等　　＋［　　］私語注意

＋［　　］分かり易さ＋［　　］有益さ

＋［　　］受講態度　＋［　　］

２）この回帰式の寄与率を求めよ。［　　］

変数自動選択で変数増減法を用いて、すべての偏回帰係数が有効である回帰モデルを作り、以下の問いに答えよ。

３）最終的な回帰式はどのようになるか。不要な変数の係数欄は空欄のままでよい。

総合評価＝　［　　］進む速さ　＋［　　］声の大きさ

＋［　　］黒板等　　＋［　　］私語注意

＋［　　］分かり易さ＋［　　］有益さ

＋［　　］受講態度　＋［　　］

４）上の回帰式の寄与率を求めよ。［　　］

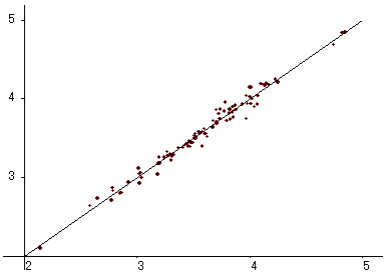
５）上の回帰式の寄与率はすべての変数を使った場合に比べ大きく下がっているか。

［大きく下がっている・あまり下がっていない］

6）この式を新しい予測モデルとして採用するか。

［採用する・採用しない］

7）予測値がどの程度実測値に近いかを見るために、下のような散布図を描け。



8）総合評価に影響を与える重要な説明変数を２つ挙げよ。

［　　］［　　］

9）データ中の最初（1 番）の授業について、総合評価の実測値，その予測値，残差（実測値と予測値の差）はいくらか。

実測値［　　］　予測値［　　］　残差［　　］

10）すべての質問項目の値が 3.5 の授業の総合評価はいくらに予測されるか。

［　　］

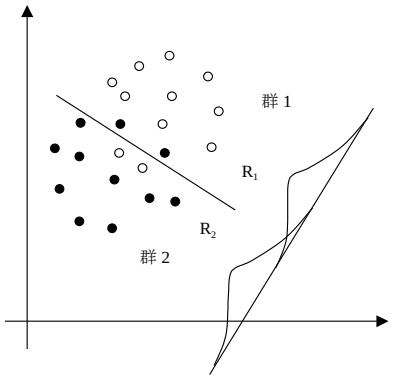
３．判別分析【第５回】

例

入学試験の合否と勉強時間・模擬試験の平均点のデータを求めたところ以下のような結果を得た（判別分析.txt (p1)）。合否を判定するための勉強時間と平均点の１次関数を求めよ。またこの関数によってこのデータを判別し、誤判別の確率を求めよ。

合否	勉強時間	平均点	合否	勉強時間	平均点
1	5.6	70.2	2	3.8	67.4
1	5.9	74.2	2	3.8	61.3
1	4.1	72.7	2	1.7	60.6
1	5.1	84.9	2	2.7	77.2
1	5.0	93.0	2	4.3	65.9
:	:	:	:	:	:
1	3.6	85.5	2	2.5	64.4
2	3.8	47.9	2	5.2	50.7
2	3.9	70.8	2	2.2	65.7

解説



判別分析の目的

2 群（多群）を判別する最適な１次式を求める。

$z = b_1$ 勉強時間 $+ b_2$ 平均点 $+ b_0$　　判別関数

判別関数の係数は？　　→　判別関数の欄

判別関数で群を分けるのは？

→　判別の分点 0（多群の場合値が最大の群）

判定に影響を与える変数は？

→　標準化係数の絶対値の大きい変数

各変数の有効性は？（要正規性・等共分散性）

→　確率の欄（係数が 0 と異なるかの検定）

誤判別の程度は？　　→　誤判別確率（実測と理論）（理論値は要正規性・等共分散性）

マハラノビス距離とは　　→　どの程度 2 群が離れているかを表わす指標

マハラノビス距離	1	4	9	16	25
誤判別確率	0.309	0.159	0.067	0.023	0.006

データ毎の判別関数の値と判別状況　　→　判別得点

問題１

判別分析.txt (p2) は、適性の有無の判定（有：1，無：2）と適性検査の結果とＳＰＩの結果を与えたデータである。判定を適性検査とＳＰＩで予測する判別分析を行い、以下の問いに答えよ。但し、事象の生起確率は各群同じ、誤判別損失は２群とも１とすること。

１）判別関数を求めよ。

判別得点＝［　　］適性検査＋［　　］ＳＰＩ＋［　　］

２）どちらの変数が判定に影響があると思われるか。［適性検査・ＳＰＩ］

３）実測値から求めた誤判別の確率は？

適性有りを無しと［　　］　適性無しを有りと［　　］

４）先頭（1 番）の人の判別得点はいくらか。［　　］

５）適性検査 50 点，ＳＰＩ 55 点の人の判別得点はいくらか、またその人の適性の有無を判定せよ。　判別得点［　　］　適性［有り・無し］

問題２

判別分析.txt (p3) はあやめの種類をがくの長さと幅、花弁の長さと幅で３群に分類したデータである。あやめの群を他の変数の１次式で判別する３群以上の判別分析を行い、以下の問題に答えよ。但し、設定は前問と同じとする。

１）３つの判別得点の式を求めよ。

判別得点１＝［　　］がくの長さ＋［　　］がくの幅
＋［　　］花弁の長さ＋［　　］花弁の幅＋［　　］

判別得点２＝［　　］がくの長さ＋［　　］がくの幅
＋［　　］花弁の長さ＋［　　］花弁の幅＋［　　］

判別得点３＝［　　］がくの長さ＋［　　］がくの幅
＋［　　］花弁の長さ＋［　　］花弁の幅＋［　　］

２）実測値から求めた誤判別確率はいくらか。

群１を他と［　　］群２を他と［　　］群３を他と［　　］

３）先頭のデータの３つの判別得点を求めよ。

判別得点１［　　］判別得点２［　　］判別得点３［　　］

これは何群に予想されたか。　［1 群・2 群・３群］

演習　判別分析【第６回】

判別分析の目的　　2 群（多群）を判別する最適な 1 次式を求める。

　　判別値＝b_1 勉強時間＋b_2 平均点＋b_0　　判別関数

判別分析が有効に利用できる条件は？　→　正規性，等共分散性（等共分散の検定）

判別関数の係数は？　　→　判別関数の欄

判別関数で群を分けるのは？　→　判別の分点 0（多群の場合は値が最大の群）

判定に影響を与える変数は？　→　標準化係数の絶対値の大きい変数

各変数の有効性は？　→　確率の欄（係数が 0 と異なるかの検定）

誤判別の程度は？　　→　誤判別確率（実測と理論）

マハラノビス距離とは　　→　どの程度 2 群が離れているかを表わす指標

データ毎の判別関数の値と判別状況　　→　判別得点

事象の生起確率とは？　→　合格・不合格の現れる確率が大きく異なっている場合の措置，各群同じかデータ数からが実用的

誤判別損失とは？　→　間違った判断をした場合の致命傷の程度

大きな差がない限り、各群 1 とするが実用的

演習 1

多変量演習 5.txt のデータを用いて、生起確率をデータ数から、誤判別損失を各群 1 として判別分析を行い、以下の問いに答えよ。合否の欄で、1 は合格、2 は不合格である。

1）判別関数を求めよ。

　判別値＝　［　　］内申＋［　　］模試 1

　　＋［　　］模試 2 ＋［　　］

2）判別の分点［　　］

3）実測値から求めた誤判別の確率は？

　合格を不合格と［　　］　不合格を合格と［　　］

4）元の設定で、各係数の有効性の検定で、5%の有意水準で有意でない変数はどれか。

　変数［　　］　検定確率［　　］

5）その変数を取り除いて再度判別分析を行い、判別関数を求めよ。但し、取り除いた変数のところは空欄とせよ。

　判別値＝　［　　］内申＋［　　］模試 1

　　＋［　　］模試 2 ＋［　　］

6）この場合、実測値から見た誤判別の確率はどうなるか。
合格を不合格と［　　］　不合格を合格と［　　］
7）元のモデルとこの新しいモデルとで誤判別確率に大きな差があると思われるか。
［大きな差がある・大した差ではない］と思われる。
8）新しいモデルで、先頭（1 番）の人の判別値はいくらか。［　　］
9）新しいモデルで、内申 3.4 点，模試１ 65 点，模試２ 70 点の人の判別値はいくらか、またその人の合否を判定せよ。
判別値［　　］　判定［合格・不合格］

演習２

多変量演習 6.txt のデータはある職業の適性について調べた結果である。適性は、1．適性あり、２．努力しだい、３．適性なしに分類され、それを予測するデータとして回答者の年齢、学力テスト、体力テスト、面接（10 段階）の結果が含まれている。

１ページ目のデータを用いて、生起確率をデータ数から、誤判別損失を各群１として判別分析を行い、以下の問いに答えよ。

1）３つの判別得点の式を求めよ。但し定数項は判別の分点を引いたものとする。
判別得点１＝［　　］年齢＋［　　］学力テスト
＋［　　］体力テスト＋［　　］面接＋［　　］
判別得点２＝［　　］年齢＋［　　］学力テスト
＋［　　］体力テスト＋［　　］面接＋［　　］
判別得点３＝［　　］年齢＋［　　］学力テスト
＋［　　］体力テスト＋［　　］面接＋［　　］
2）実測値から求めた誤判別確率はいくらか。
適性ありを他と［　　］　努力しだいを他と［　　］
適性なしを他と［　　］
3）先頭の人の３つの判別得点を求めよ。
判別得点１［　　］判別得点２［　　］判別得点３［　　］
4）先頭の人はどのように予測されているか。
［適性あり・努力しだい・適性なし］

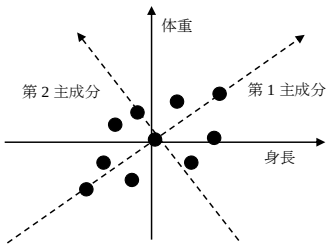
４．主成分分析【第７回】

例

以下の健康診断のデータ（Samples¥主成分分析 1.txt）から、変数の１次関数として体格を表す特徴的な指標を作り、その意味を考察せよ。

身長	体重	胸囲	座高	身長	体重	胸囲	座高
148	41	72	78	139	34	71	76
160	49	77	86	149	36	67	79
159	45	80	86	142	31	66	76
153	43	76	83	150	43	77	79
151	42	77	80	139	31	68	74
140	29	64	74	161	47	78	84
:	:	:	:	:	:	:	:
139	32	68	73	148	38	70	78

解説



主成分分析.txt（p1）のデータから、変数の１次関数として体格を表す特徴的な指標を作る。

主成分分析の目的

複数の変数を１次関数として組み合わせて、いくつかの特徴的な量を作り出す。

主成分 1 = a_{11} 身長＋a_{12} 体重＋a_{13} 胸囲＋a_{14} 座高

主成分 2 = a_{21} 身長＋a_{22} 体重＋a_{23} 胸囲＋a_{24} 座高

: :

各主成分の係数値は？　→　固有ベクトルの値（全体的に符号を変えてもよい）

各主成分のばらつき（分散）は？　→　各主成分の固有値

各主成分の重要性は？　→　各主成分の寄与率（変動の何％を表すか）

各主成分と各変数の関係は？　→　因子負荷量（各主成分と各変数の相関係数）

何番目の主成分まで意味があるか？　→　等固有値の検定（要正規性）

主成分が意味がある→他の主成分と値が異なる

データごとの主成分の値は？　→　主成分得点

共分散行列からと相関行列からどちらを使う　→　実用的には相関行列が一般的

問題

主成分分析 2.txt (p2) は生徒の教科別の成績データである。相関行列をもとにするモデルを用いて以下の問いに答えよ。

1）各主成分の固有値（分散の値)、寄与率、累積寄与率を求めよ。

	第１主成分	第２主成分	第３主成分	第４主成分	第５主成分
固有値					
寄与率					
累積寄与率					

2）主成分を２つ使うとすると第１主成分と第２主成分の関数はどのように表されるか。

第１主成分＝［　　］英語＋［　　］数学
＋［　　］国語＋［　　］理科＋［　　］社会

第２主成分＝［　　］英語＋［　　］数学
＋［　　］国語＋［　　］理科＋［　　］社会

3）これら２つの主成分で説明できるのは全体の変動の何%か。［　　］%

4）これら２つの主成分はどのように意味づけられるか。

第１主成分　意味［　　］を表す指標

第２主成分　意味［　　］を表す指標

5）先頭（1 番）の生徒の２つの主成分得点を求めよ。

第１主成分得点［　　］　第２主成分得点［　　］

6）２つの主成分の意味を考えて、この生徒にはどんな特徴があるか。

［　　］

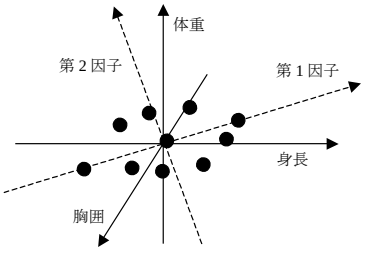
５．因子分析【第８回】

例

以下の健康診断のデータ（因子分析.txt (p1)）から、変数の背後にある体格を表す共通因子を求め、その意味を考察せよ。

身長	体重	胸囲	座高	身長	体重	胸囲	座高
148	41	72	78	139	34	71	76
160	49	77	86	149	36	67	79
159	45	80	86	142	31	66	76
153	43	76	83	150	43	77	79
151	42	77	80	139	31	68	74
:	:	:	:	:	:	:	:
139	32	68	73	148	38	70	78

解説



因子分析.txt（p1）のデータから、体格を表す共通因子を求める。

因子分析の目的

各変数の背後にある共通因子を求め、それらの１次式として各変数が表されるように係数を求める。

身長 ＝b_{11}因子 1＋b_{12}因子 2＋…

体重 ＝b_{21}因子 1＋b_{22}因子 2＋…

胸囲 ＝b_{31}因子 1＋b_{32}因子 2＋…

座高 ＝b_{41}因子 1＋b_{42}因子 2＋…　　主成分分析の逆

各因子の係数値は？　→　因子負荷量の値（全体的に符号を変えて見てもよい）

各因子と各変数の相関係数は？　→　因子負荷量の値（因子間は無相関とした場合）

各因子の重要性は？　→　各因子の寄与率

何番目の因子まで考えるか？　→　累積寄与率が 90%程度まで（寄与率も見る）

相関行列の固有値で１より大きい固有値の数

因子が各変数の変動（分散）を説明する程度は？→　共通性の値

データごとの因子の値は？　→　因子得点

問題

因子分析.txt (p2) は北海道各地の２月の気温データである。設定はデフォルトとして以下の問いに答えよ。注）江差町（えさし：南部），寿都町（すっつ：南部），小樽市（お

たる：中部），留萌市（るもい：北部），天塩町（てしお：北部）

1）各都市間の相関行列の固有値を大きい順に４つ求めよ。

1	2	3	4

以後因子数を２つと決めて各質問に答えよ。

2）各因子の寄与率と累積寄与率を求めよ。

	第１因子	第２因子
寄与率		
累積寄与率		

3）因子数は２つでよいか。［よい・注意が必要］

4）各因子の因子負荷量を求めよ。

	江差	寿都	小樽	留萌	天塩
第１因子					
第２因子					

5）上の因子負荷量の値から各因子の意味を解釈せよ。

第１因子：［　　］の気温を代表する因子

第２因子：［　　］の気温を代表する因子

6）各地の気温の変動は因子によりどの程度説明されるか（共通性）。

江差	寿都	小樽	留萌	天塩

7）最初の３日間の因子の値（因子得点）を推定せよ。

	第１因子	第２因子
1		
2		
3		

8）この３日間、北海道はどのような気候だったか。

［　　］

9）このモデルは良いモデルと思うか。

［良いと思う・あまり良いと思わない］

演習【第９回】

多変量演習 8.txt のデータはある学校で測定した小学６年生の運動適性テストの結果である。因子分析を用いて特徴を分析し、以下の問いに答えよ。

１）各科目間の相関行列の固有値を大きい順に求めよ。

1	2	3	4	5

２）因子数を２つとして、因子分析を行い、寄与率を求めよ。

因子１	因子２

３）因子数を３つとして、因子分析を行い、寄与率を求めよ。

因子１	因子２	因子３

本来なら因子数３が良いが、解釈の問題から、以後因子数を２つと決めて各質問に答えよ。但し、バリマックス回転ありとすること

４）各因子の因子負荷量を求めよ。

	立幅跳び	腹筋	腕立伏せ	往復走	５分間走
第１因子					
第２因子					

５）この場合の各因子の意味を解釈せよ。

第１因子：[　　]を表す因子

第２因子：[　　]を表す因子

6）先頭から３人及び、特徴的な９番目の人の因子の値（因子得点）を推定せよ。

	第１因子	第２因子
1		
2		
3		
9		

7）９番目の人にはどんな特徴があるか。因子負荷量の符号に注意して考えよ。

［　　］

8）各種目の変動は因子によりどの程度説明されるか。

立幅跳び	腹筋	腕立伏せ	往復走	５分間走

9）予測値が２つの因子から予測されたことを考えると、この分析はうまくいったと思うか。　［まずまずうまくいった・うまくいっていない］

６．クラスター分析【第 10 回】

例

表　各人の好みを 1～9 の点数で表わした表（Samples¥クラスター分析 1.txt）

	日本酒	焼酎	ビール	ウィスキー	ワイン
増川	1	2	9	6	5
西山	3	1	7	5	4
三好	5	3	4	2	2
芝田	3	6	2	8	3
尾崎	4	6	9	3	4
藤田	7	2	5	4	5
細川	7	5	4	3	2

クラスター分析の目的

１）回答の類似度で個人を分類する。　→　個体（レコード）の分類

２）回答の類似度で変数を分類する。　　→　変数の分類

クラスター分析は分類をどのように表示するか　→　デンドログラム（解答参照）

デンドログラムの縦軸は　→　要素またはクラスター間の距離（類似の程度を示す量）

要素間の距離とは

個体間について

量的データ：ユークリッド距離、標準化ユークリッド距離、マハラノビス距離等

質的 0/1 データ：類似比、一致係数、φ係数等を使ったもの

変数間について

量的データ：1-相関係数、1-相関係数|、1-順位相関係数、1-|順位相関係数|

質的データ：平均平方根一致係数、一致係数、クラメールのＶ等を使ったもの

要素間の距離を知るには　→　距離行列

クラスター構成でよく使われる方法　→　最長距離法、ウォード法

クラスター構成過程を表示するには　→　クラスター構成と距離

問題１

Samples¥クラスター分析 4.txt はある野球チームの今年度の成績である。これについてクラスター分析を行い以下の問いに答えよ。

１）ユークリッド距離及び標準化ユークリッド距離を用いた場合、山下と田中の距離はいくらか。ユークリッド距離［　　］　標準化ユークリッド距離［　　］

2）各変数の標準偏差はいくらか。

打率	安打	本塁打	打点	盗塁

3）上の結果から、距離測定法はどちらを利用すべきか。

［ユークリッド距離・標準化ユークリッド距離］　以後はこの距離を用いる。

4）クラスター構成法を最長距離法とする場合、最初にクラスターを構成するのはどの要素とどの要素でそれらの距離はいくらか。

［　　］と［　　］で距離［　　］

5）最長距離法の場合、４分類か５分類が適当と思われるが、４分類の場合、各クラスターにはどのような要素が含まれるか。

［　　］［　　］［　　］［　　］

6）最長距離法と最短距離法とでどちらの分類が理解しやすいと思われるか。

［最長距離法・最短距離法］

8）１－相関係数の距離測定法で最長距離法を用いて変数を３分類すると各クラスターに含まれる要素はどのようになるか。

［　　］［　　］［　　］

問題２

Samples¥クラスター分析 3.txt のデータを用いてクラスター分析を行い、以下の問いに答えよ。

1）個体の分類

距離測定法は標準化ユークリッド距離、クラスター構成法は最長距離法を用いると、３分類の場合、各クラスターに含まれる要素はどうなるか。

［　　］［　　］［　　］

2）変数の分類

距離測定法は１－相関係数、クラスター構成法は最長距離法を用いると、２分類の場合、各クラスターに含まれる要素はどうなるか。

［　　］［　　］

演習　クラスター分析【第 11 回】

クラスター分析の目的

　１）類似度による個体（レコード）の分類

　２）類似度による変数の分類

クラスター分析は分類をどのように表示するか　→　デンドログラム

デンドログラムの縦軸は　→　要素またはクラスター間の距離（類似の程度を示す量）

要素間の距離とは

　個体間について

　　量的データ：ユークリッド距離、標準化ユークリッド距離、マハラノビス距離等

　　質的 0/1 データ：類似比、一致係数、ϕ 係数等を使ったもの

　変数間について

　　量的データ：相関係数、順位相関係数等を使ったもの

　　質的データ：平均平方根一致係数、一致係数、クラメールのＶ等を使ったもの

要素間の距離を知るには　→　距離行列

クラスター構成法

　最短距離法（棒状の分布に最適）

　最長距離法（クラスターを分離する能力が高い）

　他に、群平均法、重心法、メジアン法、ウォード法

クラスター構成過程を表示するには　→　クラスター構成と距離

演習 1

　多変量演習 9.txt は学生による授業評価のデータであり、レコード（個体）は１つの授業で調べた質問項目（変数）ごとの平均を表している。このデータからクラスター分析を用いて、個体や変数の類似性の特徴を見出したい。以下の質問に答えよ。

１）ユークリッド距離を用いた場合、1 番と 12 番の距離はいくらか。［　　］

２）クラスター構成法を最長距離法、距離測定法をユークリッド距離とする場合、最初にクラスターを構成するのは何番と何番でそれらの距離はいくらか。

　個体［　　］番と個体［　　］番で、距離［　　］

３）上の設定で、最初にクラスターとクラスター、またはクラスターと要素の結合になるのはどのようなクラスター（要素）か。それらに含まれる要素を示せ。またその際の距離はいくらか。

　クラスター［　　］とクラスター（要素）［　　］　距離［　　］

4）上の設定でクラスター分析を実行し、4つのクラスターに分けたとき、それらのクラスターに含まれる要素（授業の番号）は何か。

［　　］［　　］［　　］［　　］

5）5番が含まれるクラスターと10番が含まれるクラスターの最も大きな特徴は何か。

5番［　　］　10番［　　］

6）距離測定法を標準化ユークリッド距離（各変数を標準化したときのユークリッド距離）に変えた場合、クラスター構成は大きく変わるか。

［変わる・あまり変わらない］　　注）標準化値＝（値－平均値）／標準偏差

7）これにはどんな理由が考えられるか。

各変数の［　　］があまり違わないから。

8）距離測定法をユークリッド距離とし、クラスター構成法を最短距離法に変えるとクラスター構成は大きく変わるか。［変わる・あまり変わらない］

9）ユークリッド距離の場合、その他のクラスター構成法は最長距離法と最短距離法のどちらに近いか。［最長距離法・最短距離法］

各質問についての分類を行いたいが、距離測定法を1－相関係数として以下の問いに答えよ。

10）最長距離法で上の距離測定法を用いる場合、最初にクラスターを構成するのは何と何で、そのときの距離はいくらか。

変数［　　］と変数［　　］で、距離［　　］

11）上の設定でクラスター分析を行い、変数を3つのクラスターに分類する場合、それらのクラスターに含まれる要素（変数）は何か。

［　　］［　　］［　　］

７．正準相関分析【第 12 回】

例

正準相関分析 1.txt のデータを用いて、複数の変数間で相関の高い特徴的な量を求める。

身長	座高	体重	胸囲
148	78	41	72
160	86	49	77
159	86	45	80
153	83	43	76
:	:	:	:
148	78	38	70

正準相関分析の目的

複数の変数から作られる２つの群の中で特徴的な量を見出し、それらの最大の相関を求める。

どのようにして相関を考えるのか。

$y = a_1$身長$+ a_2$座高
$z = b_1$体重$+ b_2$胸囲　　正準変数の組 y と z が最大の相関を持つよう係数を選ぶ。

y と z の最大の相関とは　→　正準相関係数（変数の組によって複数ある）

係数はどのように表示されるか。　→　正準相関分析で正準変数１係数と同２係数

正準変数 y と z の各データの値を見るには　→　正準変数値

各変数と同じ群の正準変数との関係は　→　正準負荷量（相関係数）、解釈に利用

各変数と違う群の正準変数との関係は　→　交差負荷量（相関係数）、解釈に利用

複数の正準変数の組が得られるが、他の正準変数の組同士の関係は　→　相関係数 0

問題

正準相関分析 2.txt について、文系科目（英語・国語・社会）と理系科目（数学・理科）に分け、正準相関分析を実行し、以下の問いに答えよ。但し、相関行列を用いたモデルで、第１正準変数について考えること。

１）文系科目と理系科目の正準相関係数はいくらか。［　　］

２）文系科目と理系科目の正準変数はそれぞれどのように表されるか。

文系正準変数＝［　　］英語＋［　　］国語＋［　　］社会

理系正準変数＝［　　］数学＋［　　］理科

3）各変数の正準負荷量の値はいくらか。

英語	国語	社会	数学	理科

4）各変数の交差負荷量の値はいくらか。

数学	理科	英語	国語	社会

5）各正準変数と最も相関のある同じ組の科目は何か。
文系正準変数では［英語・国語・社会］、理系正準変数では［数学・理科］

6）各正準変数と最も相関のある違う組の科目は何か。
文系正準変数へは［数学・理科］、理系正準変数へは［英語・国語・社会］

7）各科目の平均と標準偏差（不偏分散からのもの）を求め、
標準化変数＝（値－平均）／標準偏差
の式によって、英語 60、国語 72、社会 66、数学 58、理科 55 の人の標準化変数値を求めよ。

科目	英語	国語	社会	数学	理科
標準化変数値					

8）上の標準化値を利用して、この人の正準変数の値を求めよ。
文系正準変数［　　］　理系正準変数［　　］

８．数量化Ｉ類【第 13 回】

例

以下の地域（1：都市部、2：山村部)、気候（1：温暖、2：平均的、3：寒冷)、ある商品の販売率のデータ（数量化Ｉ類 1.txt）から販売率（目的変数）を予測する式を作り、それがどの程度有効か検討する。

販売率	地域	気候
3.0	1	2
1.8	2	1
:	:	:
2.3	1	3

販売率	地域-1	地域-2	気候-1	気候-2	気候-3
3.0	1	0	0	1	0
1.8	0	1	1	0	0
:	:	:	:	:	:
2.3	1	0	0	0	1

左のようなアイテムのデータから、それぞれのアイテムが複数のカテゴリに分かれる右の形のデータを作る。

このデータをもとに以下の式で目的変数を予測する。

$$Y = a_{11}x_{11} + a_{12}x_{12} + a_{21}x_{21} + a_{22}x_{22} + a_{23}x_{23} + a_{00}$$

カテゴリウェイト、重回帰カテギリウェイト、基準化カテゴリウェイトの違いは

→　予測値を計算する上では同じ（予測値への影響の見易さが異なる）

予測値と実測値との相関係数　→　重相関係数

予測値は実測値をどれだけ説明しているか　→　寄与率

各アイテムの重要性は　→　アイテム重要性ボタンのウェイト範囲、p 値

予測値と実測値の散布図　→　散布図ボタン

問題

数量化Ｉ類 2.txt は店舗の売り上げを立地、人通り、競合の３段階分類データで予測しようとするものである。

１）カテゴリウェイト（定数項を０）を用いた予測式を表せ。

予測売り上げ＝［　　］立地１＋［　　］立地２＋［　　］立地３
＋［　　］人通り１＋［　　］人通り２＋［　　］人通り３
＋［　　］競合１＋［　　］競合２＋［　　］競合３＋［　　］

２）重回帰カテゴリウェイト（各先頭アイテムを基準）を用いた予測式を表せ。

予測売り上げ＝［　　］立地１＋［　　］立地２＋［　　］立地３
＋［　　］人通り１＋［　　］人通り２＋［　　］人通り３
＋［　　］競合１＋［　　］競合２＋［　　］競合３＋［　　］

3）基準化カテゴリウェイトを用いた（目的変数の平均値を基準）予測式を表せ。
予測売り上げ＝［　　］立地１＋［　　］立地２＋［　　］立地３
＋［　　］人通り１＋［　　］人通り２＋［　　］人通り３
＋［　　］競合１＋［　　］競合２＋［　　］競合３＋［　　］

4）各アイテムで予測売り上げを増やす方に寄与するアイテムはどれか。すべてに○を付けよ。
［立地１・立地２・立地３］　［人通り１・人通り２・人通り３］
［競合１・競合２・競合３］

5）予測式は実測値の変動を何%予測できるか。［　　］%

6）立地：２，人通り：２，競合：２の店舗の売り上げを予測せよ。［　　］

7）ウェイト範囲で見る場合、予測値に最も大きな影響を与えるアイテムは何か。
［立地・人通り・競合］

8）数量化Ｉ類と同じ分析を 0/1 データを用いた重回帰分析で行った。但し、各アイテムの第１カテゴリは係数が０として､変数から外した｡そのときの重回帰式を示せ。
予測売り上げ＝［　　］立地２＋［　　］立地３
＋［　　］人通り２＋［　　］人通り３
＋［　　］競合２＋［　　］競合３＋［　　］

9）このことから上の重回帰分析と数量化Ｉ類は［同じ・異なる］ものと考えられる。

この方法を利用して、重回帰分析の説明変数の中に質的データを入れることができる。また、同様にして判別分析の説明変数の中に質的データを入れることもできる。

９．ロジスティック回帰分析【第 14 回】

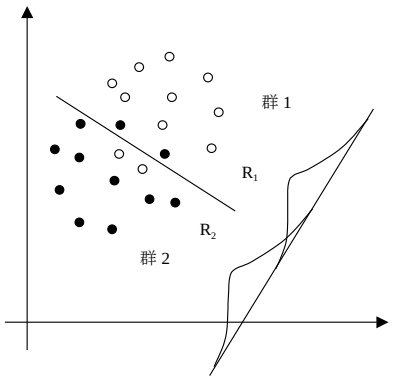
事象１か事象２か、２つの事象のうちどちらが発生するかを予想する問題を考える。判別関数値の正負を用いてこの判定を行うのが判別分析であるが、事象１の起きる確率を直接求める分析がロジスティック回帰分析である。

例

入学試験の合否と勉強時間・模擬試験の平均点のデータを求めたところ以下のような結果を得た（Samples¥判別分析１.txt)。合格確率を求めるための勉強時間と平均点の１次式を求めよ。またこの式によってこのデータを判別し、誤判別の確率を求めよ。

合否	勉強時間	平均点	合否	勉強時間	平均点
1	5.6	70.2	2	3.8	67.4
1	5.9	74.2	2	3.8	61.3
1	4.1	72.7	2	1.7	60.6
1	5.1	84.9	2	2.7	77.2
1	5.0	93.0	2	4.3	65.9
:	:	:	:	:	:
1	3.6	85.5	2	2.5	64.4
2	3.8	47.9	2	5.2	50.7
2	3.9	70.8	2	2.2	65.7

解説



ロジスティック回帰分析の目的

2 群（多群）の１つの事象の発生確率 p を（対数オッズの形で）推定する最適な１次式を求める。

$$\log\frac{p}{1-p} = b_1 \text{勉強時間} + b_2 \text{平均点} + b_0$$

回帰式の係数は？　→　偏回帰係数の欄

判別の推定で群を分けるのは？

→　確率（事象の発生確率）の値 0.5

確率推定に影響を与える変数は？

→　標準化値の絶対値の大きい変数

各係数の有効性は？　→　両側確率の欄（係数が 0 と異なるかの検定）

誤判別の程度は？　→　誤判別確率

分析の精度は？　→　逸脱度、尤度比

係数変化による事象の発生オッズ比は？　→　EXP(b)の欄

データ毎の予測確率の値と判別状況　→　「予測確率と予測値」

問題

ロジスティック回帰.txt (p2) のデータを用いて以下の問いに答えよ。データは、先頭列で群分け形式で、対象を 1（発症群)、モデルはロジスティックモデルとすること。

1)対数オッズを予測する回帰式の偏回帰係数の値を求めよ。(p は予測発症確率である)

$$\log\frac{p}{1-p} = [\quad\quad] \text{要因1} + [\quad\quad] \text{要因2} + [\quad\quad]$$

2）これら 3 つの係数は 0 でないといえるか。

要因 1 係数　検定確率［　　］　0 と異なると［いえる・いえない］

要因 2 係数　検定確率［　　］　0 と異なると［いえる・いえない］

切片　検定確率［　　］　0 と異なると［いえる・いえない］

3）各要因の有無による発症オッズの比（罹患危険率の比）は EXP(b)の欄で与えられているが、2 つの要因でそれぞれいくらか。

要因 1［　　］　要因 2［　　］

4）最適値からのずれを表す逸脱度、最小モデルからのずれを表す尤度比の値はいくらか。これらの値から、このモデルは有効と考えられるか。

逸脱度［　　］　モデルは［有効・有効でない］

尤度比［　　］　モデルは［有効・有効でない］

注）逸脱度は小さいほど良い（p>0.05)、尤度比は大きいほど良い（p<0.05)。

5）所属群の判定で、誤判別確率はいくらか。

1 群（合格群）を他と　［　　］

0 群（不合格群）を他と［　　］

6）判別の分点は予測確率がいくらのところか。［　　］

7）4 番目の人の実測値、予測確率、予測値を求めよ。

実測値［　　］，予測確率［　　］，予測値［　　］