

合否判定できない判別分析の総括

— 分散共分散行列によるLDFとQDFの使命の終焉 —

成蹊大学 経済学部

新村秀一

分散共分散による判別分析の問題

- 2群が多次元正規分布し分散共分散行列が等しいと仮定すれば、分散比最大化基準によるLDFが、容易に2群を表す正規分布 $N(m_1, s_1)$ と $N(m_2, s_2)$ の対数尤度で定式化。
- 2群の分散共分散行列による判別手法
 - 2群の分散共分散行列が等しい場合, $LDF: f(x) = \{x - (m_1 + m_2)/2\}' \Sigma^{-1} (m_1 - m_2)$
 - 2群の分散共分散行列が等しくない場合, 2次判別関数(QDF)
 $f(x) = x' (\Sigma_2^{-1} - \Sigma_1^{-1}) x / 2 + (m_1' \Sigma_1^{-1} - m_2' \Sigma_2^{-1}) x + c$
 - マハラノビスの汎距離から, 多群判別や品質管理のMT理論
 $D = \text{SQRT} \left((x - m)' \Sigma^{-1} (x - m) \right)$
この式に重大な問題が見過ごされてきた!
- 試験の合否判定を, 得点を説明変数としてQDFで判別すると, 合格群の全てが不合格群に誤判別される理由が解明できた
- 試験の合否判定を大問4問で行い合格最低点を50点 $F = T1 + T2 + T3 + T4 - 49.5$ で, $f > 0$ なら合格, $f < 0$ なら不合格

1. はじめに

- 判別分析は, Fisher[1]が2群の分散比の最大化からLDF(線形判別関数)を定式化したが, 正規分布の対数尤度から同じLDFがスマートに再定義される.
 - 統計ソフトに取り入れやすい
 - 分散共分散行列から, LDFやQDFさらにマハラノビスの汎距離を用いた多群判別. 品質管理のMT理論. ゲノム判別
 - 線形分離可能なデータを認識できない.
 - 合否判定できないことはすぐに確認できる.
- 判別規則の単純さに隠れて多くの問題が隠蔽
 - $y_i * f(x) > 0 \rightarrow \text{class1/class2}$ に判別, $y_i * f(x) < 0 \rightarrow \text{class1/class2}$ に誤判別
 - 判別境界上のケースの扱いは未解決.
 - MNM (Minimum Number of Misclassification) 基準による最適線形判別関数(OLDF)で解消.

発表の概要

- 2010年から2012年まで3年間行っていた「統計入門」の中間と期末試験(10択100問の試験, 4個の大問に分類)の総合報告
 - 統計家は, 大学の試験のデータを分析し, FDに貢献できる
- 合格得点の3水準(10%, 50%, 90%点)で合否判定
 - 大問の合否判定を, OLDF, ロジスティック回帰, LDF, QDF, SVMで行う
 - LDFとQDFは, 合否判定できない
 - LDFの誤分類確率は[2.3, 16.7], QDFは[0.8, 10.8]
 - QDFは, 小問の合否判定で合格群すべてが不合格群に誤判別される理由

2. 単純な判別規則と判別分析の問題点

(1) Fisherの仮説の問題

- かつてはFisherの仮説を満たさないデータにLDFを適用してはいけないという研究者
- 多次元正規性の検定はできていない
- 現実のデータはこの仮説を満たすものは少ない
- 医学診断で群の平均は典型症例でない

(2) 判別超平面上のケースの帰属

- $f(x_i)=0$ のケースをどちらに判別するかは未解決

(3) 標本誤分類確率と母誤分類確率の関係

- Miyake & Shinmura[18]参照

(4) 3つの判別境界と誤分類数の問題

- 判別境界は3つの異なった決め方
 - 基本は、2群が正規分布 $N(m_1, s_1)$ と $N(m_2, s_2)$ と考えて対数尤度 $(\log(N(m_1, s_1)/N(m_2, s_2)))$ が0になる判別境界.
- しかし判別境界を動かすと、得られた誤分類数より小さなものが簡単に得られることが多い.
- 事前確率とリスク概念で、正規分布を修正.
 - ケース数 (n_1, n_2) に比例させた事前確率で対数尤度を修正: $\log(n_1 \times N(m_1, s_1)/(n_2 \times N(m_2, s_2)))$.
 - 医学診断で異常群を正常群に間違う危険性を勘案し、リスク (r_1, r_2) で修正:
 - $(\log(r_1 \times n_1 \times N(m_1, s_1)/(r_2 \times n_2 \times N(m_2, s_2))))$.
- 正規分布を事前確率で修正したものを基本とすべき

(5) MNMの正当性

- 135個の異なった判別モデルの100重交差検証法
 - LDFは120個,
 - ロジスティック回帰は102個の平均誤分類確率が改定IP-OLDFより悪い[12].

(6) MNM=0を認識できない問題点

- 線形分離可能という専門用語が統計理論にない
- LDFやQDFはMNM=0の空間を認識できない

(7) 誤分類数と判別係数の95%信頼区間

- 判別係数は定数項が正と0と負の3つの異なった構造
- MNMが最少な最適凸体の内点を判別係数とすれば、判別分析の問題が解明

(8) 合否判定できない問題

3. 試験の合否判定

- 試験の合否判定は、**自明な線形分離可能な判別が可能**
 - 50点以上を合格とする場合: $y=T1+T2+T3+T4-49.5$ で $y>0$ であれば合格, $y<0$ であれば不合格
 - しかし, LDFやQDFは合否判定できない
- 誰もがすぐに手に入るMNM=0の良質な研究データ
- 大学の統計研究者は、積極的に試験データの統計分析を行うことで、FD活動に貢献できる[13].
 - 統計入門で、正規分布表が意外と新入生に難しい
 - 大問で変数選択を行えば、設問の難易度がある程度分かる

3.1 授業の概要

| 週 | 2010年(2012年度) | 2011年 |
|----|-------------------|-------|
| 1 | PowerPointで概論 | 同左 |
| 2 | 最頻値, 中央値, 平均値 | 同左 |
| 3 | 範囲, 四分位範囲, SD, CV | 同左 |
| 4 | 学生データの解釈 | 同左 |
| 5 | 正規分布 | 同左 |
| 6 | 自由度, SE, t分布 | 相関係数 |
| 7 | 中間試験 | 中間試験 |
| 8 | 相関係数 | 9回目 |
| 9 | Excelで相関の計算 | 10回目 |
| 10 | 単回帰分析 | 12回目 |
| 11 | 単回帰分析 | 期末試験 |
| 12 | 分割表と独立性の検定 | |
| 13 | 分割表と独立性の検定 | |
| 14 | まとめ | |
| 15 | 期末試験 | |

- 統計の入門科目として, 基本統計量, 相関, 単回帰, 分割表
- 4件のデータで統計量を説明, Excelで相関と単回帰分析の計算, JMPで実際の出力の解釈
- 2011年は, 電力節減のため11週
- 試験は10択100問のマークセンス試験
- 試験実施後, 得点と統計分析した内容を学生にフィードバック

表2 4個の大問

| 大問 | 中間試験 | | 期末試験 | |
|----|--------|----|--------|----|
| | 内容 | 得点 | 内容 | 得点 |
| T1 | 基礎統計量 | 29 | 統計量の筆算 | 26 |
| T2 | 統計量の筆算 | 12 | 相関と回帰 | 30 |
| T3 | 正規分布 | 19 | 分割表 | 21 |
| T4 | JMPの解釈 | 40 | JMPの解釈 | 23 |

- 100個の小問と4個の大問で, 試験の質の評価を行う.
- 実際の合格最低点は10%点であるが, 50%点と90%点で継続分析
- 各水準ごとに変数選択法とMNM=0になる最小設問を調べることで, 設問の難易度が分かる

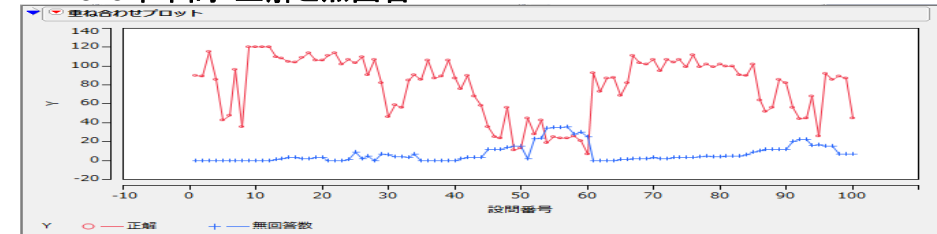
3.2 2012年の欠席者増大の影響の分析

(1) 3年間の成績評価

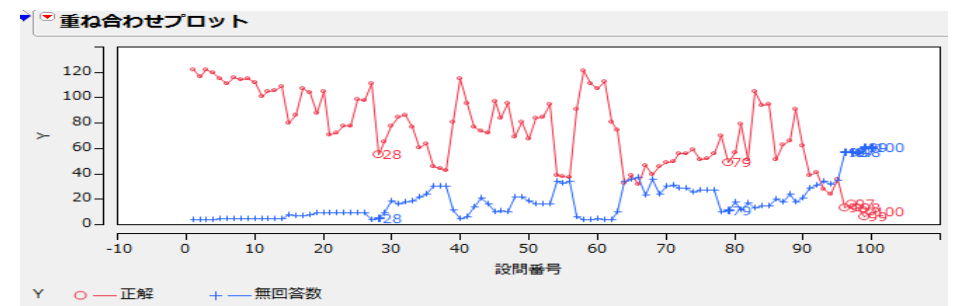
- 2年間の経験を踏まえ, 成績の上昇を期待
- 中間試験以降欠席が増える
 - 140人中, 欠席者が40人から60人に増えつづける
 - 例年は, 中間試験後に40人に増え, 減っていく
 - 得点分布が2峰性に?
- 結論
 - 2010年より悪い
 - 相関, 単回帰, 分割表より

| | 2010 | 2011 | 2012 |
|------|---------------|------|------|
| 0% | 31 | 25 | 21 |
| 10% | 48 | 42 | 37 |
| 中間 | 50% 66 | 61 | 63 |
| 90% | 82 | 79 | 78 |
| 100% | 93 | 88 | 88 |
| 平均 | 65.1 | 56.1 | 58.8 |
| 0% | 22 | 26 | 20 |
| 10% | 40 | 43 | 41 |
| 期末 | 50% 60 | 60 | 58 |
| 90% | 82 | 81 | 81 |
| 100% | 91 | 99 | 95 |
| 平均 | 59.3 | 57.1 | 58.8 |
| r | 0.54 | 0.7 | 0.51 |
| R2 | 0.29 | 0.49 | 0.26 |

2010年中間: 正解と無回答



2010年期末: 正解と無回答



2010年度のグラフ

評価1(中間+期末=200点)と
評価2(中間+期末+宿題=230点)を
5段階にした分割

中間と期末の散布図
未受験者と、得点変動の激しい学生

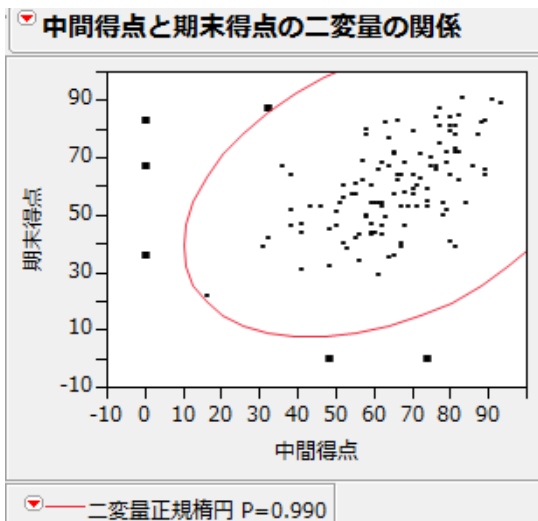
分割表

| | | 評価2 | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--|
| 度数 | 1 | 2 | 3 | 4 | 5 | | |
| 期待値 | 17 | 4 | 0 | 0 | 0 | 21 | |
| 偏差 | 3.024 | 6.216 | 6.384 | 3.36 | 2.016 | | |
| | 13.976 | -2.216 | -6.384 | -3.36 | -2.016 | | |
| 評価1 | 1 | 1 | 30 | 6 | 0 | 37 | |
| | 2 | 5.328 | 10.952 | 11.248 | 5.92 | 3.552 | |
| | 3 | -4.328 | 19.048 | -5.248 | -5.92 | -3.552 | |
| | 4 | 5.184 | 10.656 | 10.944 | 5.76 | 3.456 | |
| | 5 | -5.184 | -7.656 | 17.056 | -0.76 | -3.456 | |
| 評価2 | 1 | 0 | 3 | 28 | 5 | 36 | |
| | 2 | 2.88 | 5.92 | 6.08 | 3.2 | 1.92 | |
| | 3 | -2.88 | -5.92 | -2.08 | 7.8 | 3.08 | |
| | 4 | 0 | 0 | 4 | 11 | 5 | |
| | 5 | 1.584 | 3.256 | 3.344 | 1.76 | 1.056 | |
| | -1.584 | -3.256 | -3.344 | 2.24 | 5.944 | | |
| | 18 | 37 | 38 | 20 | 12 | 125 | |

検定

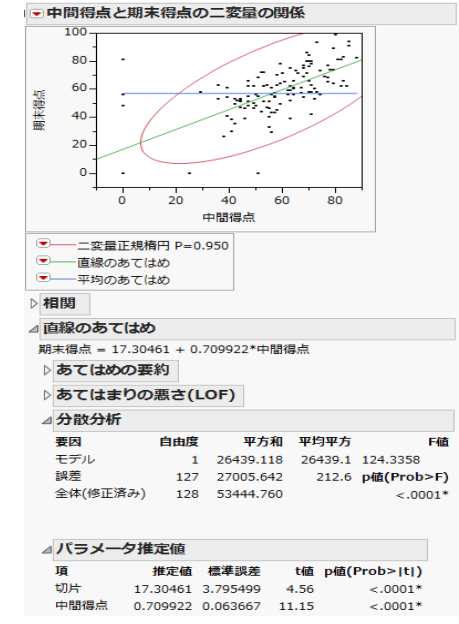
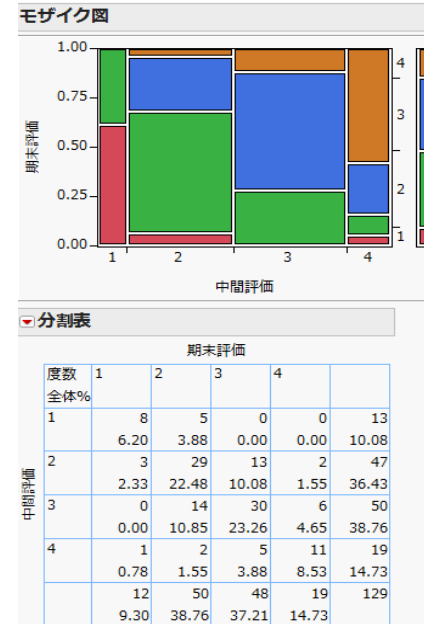
| N | 自由度 | (-1)*対数尤度 | R2乗(U) |
|-----|-----|-----------|--------|
| 125 | 16 | 107.38645 | 0.5654 |

検定 カイ2乗 p値(Prob>ChiSq)
尤度比 214.773 <.0001*
Pearson 244.487 <.0001*



2011年の分割表: 上位は1ランク落ち

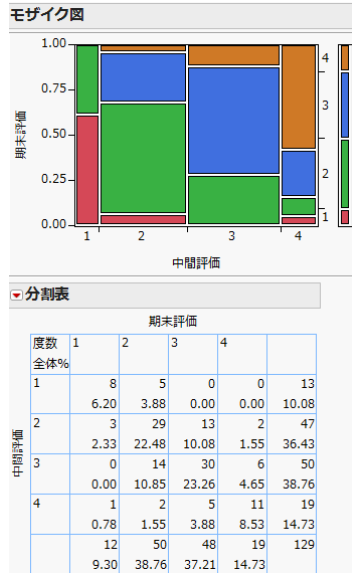
散布図: 未受験以外外れ値なし



(3)分割表による評価の変動の分析

2011年: 対角線上が多い

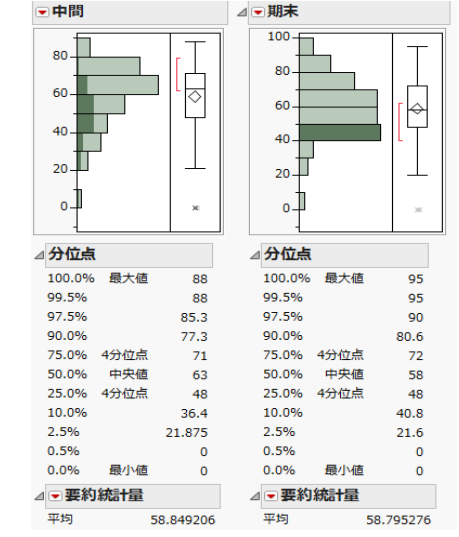
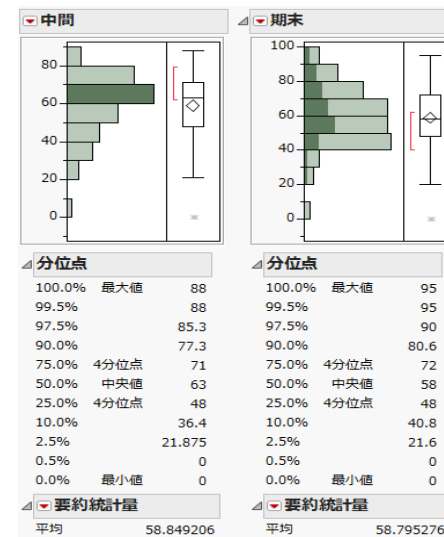
2012年: 上位からの転落



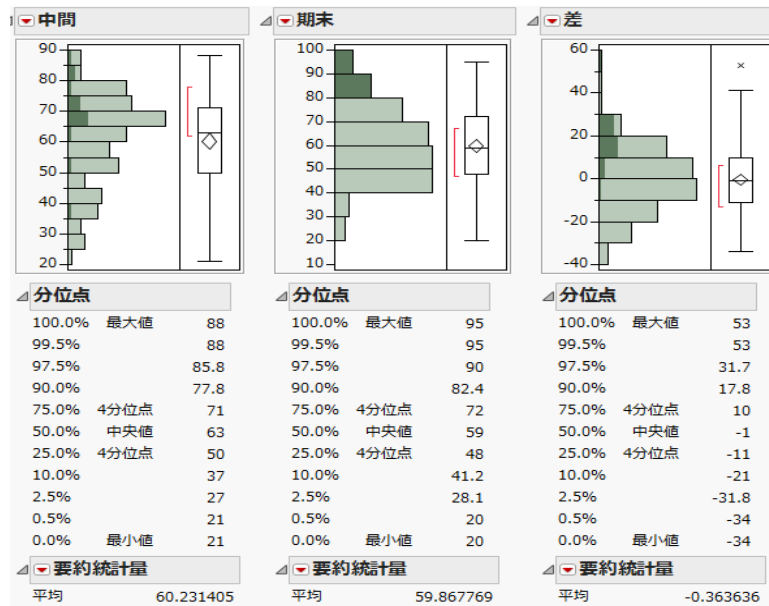
2012年度中間と期末の得点分布

中間の最頻値

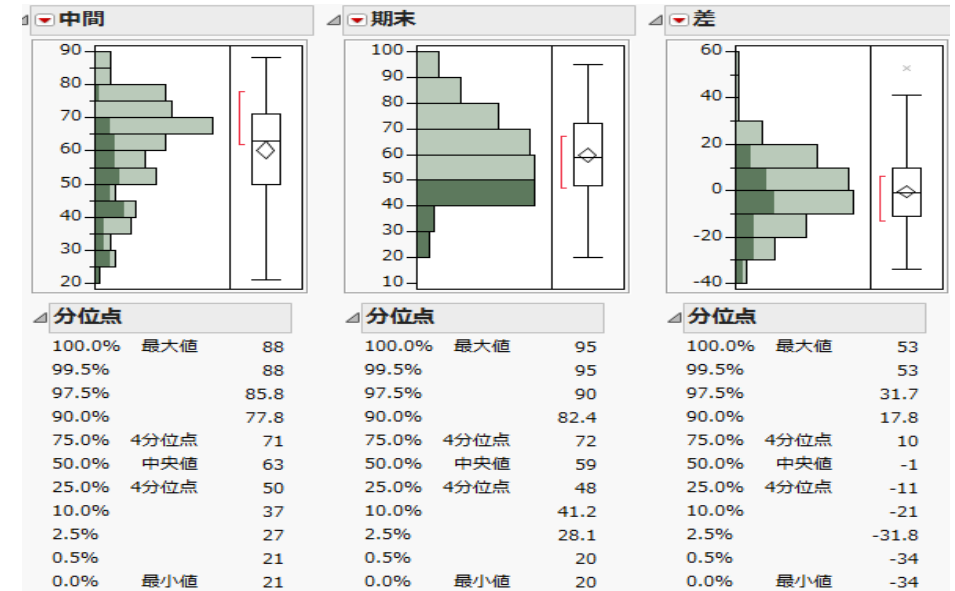
期末の最頻値



2012年の期末成績上位



2012年の期末成績下位群



4. 大問と小問による合否判定

- 10択100問の小問を, 4個の大問にまとめる
 - 中間の計算は容易である. 正規分布が難しいことが分かる
 - 期末の計算は難しい
 - 2010年と2011年は分割表が時間不足で未消化
- 研究では, 10%点, 50%点, 90%点で合否判定

| 大問 | 中間試験 | | | 期末試験 | | |
|----|--------|----|------------|--------|----|--------|
| | 内容 | 得点 | 小問番号 | 内容 | 得点 | 小問番号 |
| T1 | 基礎統計量 | 29 | 1-8, 21-41 | 計算 | 26 | 1-26 |
| T2 | 計算 | 12 | 9-20 | 相関と回帰 | 30 | 27-56 |
| T3 | 正規分布 | 19 | 42-60 | 分割表 | 21 | 57-77 |
| T4 | JMPの解釈 | 40 | 61-100 | JMPの解釈 | 23 | 78-100 |

4.1 大問の分析

- 3水準の合否判定で, 大問の難易度と合否判定に必要/不要がある程度説明可
- 最適線形判別関数とロジスティック回帰が合否判定できる次元で, LDFとQDFは合否判定できない.
- 2012年の10%点のQDF以外, LDFとQDFは合否判定できない.

| | | 10% | | | | | 50% | | | | | 90% | | | | |
|----|------|-----|-----|------|-----|----|-----|-----|------|-----|----|-----|-----|------|-----|----|
| | | P | MNM | Logi | LDF | QD | P | MNM | Logi | LDF | QD | P | MNM | Logi | LDF | QD |
| 中間 | 2010 | 4 | 0 | 0 | 9 | 2 | 4 | 0 | 0 | 3 | 6 | 3 | 0 | 0 | 20 | 10 |
| | 2011 | 3 | 0 | 0 | 9 | 10 | 4 | 0 | 0 | 3 | 3 | 3 | 0 | 0 | 13 | 5 |
| | 2012 | 2 | 0 | 0 | 1 | 1 | 4 | 0 | 0 | 7 | 5 | 4 | 0 | 0 | 10 | 3 |
| 期末 | 2010 | 4 | 0 | 0 | 5 | 2 | 4 | 0 | 0 | 4 | 5 | 4 | 0 | 0 | 4 | 13 |
| | 2011 | 4 | 0 | 0 | 16 | 4 | 4 | 0 | 0 | 4 | 5 | 4 | 0 | 0 | 5 | 12 |
| | 2012 | 4 | 0 | 0 | 9 | 3 | 4 | 0 | 0 | 3 | 3 | 4 | 0 | 0 | 4 | 1 |

中間の大問の分析 (上:2010, 中:2011, 下:2012)

10%点: T3の正規分布が難しい, 2012年はT4のJMPの解釈だけで合否判定
 50%点: 2010年と2011年はT2の計算は不要, 2012年はT3の正規分布は不要
 90%点: 2010年と2011年はT2の計算とT1の基礎統計量は不要

| P | Var | MNM | Logi | LDF | QDF | Var | MNM | Logi | LDF | QDF | Var | MNM | Logi | LDF | QDF |
|---|-----|-----|------|-----|-----|-----|-----|------|-----|-----|-----|-----|------|-----|-----|
| 1 | T4 | 6 | 9 | 11 | 11 | T4 | 16 | 16 | 16 | 16 | T3 | 10 | 27 | 24 | 24 |
| 2 | T2 | 2 | 6 | 11 | 9 | T3 | 9 | 10 | 12 | 12 | T4 | 5 | 10 | 20 | 11 |
| 3 | T1 | 1 | 3 | 8 | 5 | T1 | 2 | 2 | 5 | 6 | T1 | 0 | 0 | 20 | 10 |
| 4 | T3 | 0 | 0 | 9 | 2 | T2 | 0 | 0 | 3 | 6 | T2 | 0 | 0 | 20 | 11 |
| 1 | T2 | 9 | 17 | 15 | 15 | T4 | 9 | 9 | 9 | 9 | T3 | 6 | 7 | 14 | 14 |
| 2 | T4 | 4 | 9 | 11 | 9 | T1 | 4 | 4 | 5 | 7 | T4 | 1 | 1 | 14 | 6 |
| 3 | T1 | 0 | 0 | 9 | 10 | T3 | 1 | 2 | 3 | 3 | T1 | 0 | 0 | 13 | 5 |
| 4 | T3 | 0 | 0 | 9 | 11 | T2 | 0 | 0 | 3 | 3 | T2 | 0 | 0 | 14 | 9 |
| 1 | T4 | 4 | 8 | 6 | 6 | T4 | 12 | 12 | 14 | 12 | T3 | 8 | 30 | 12 | 12 |
| 2 | T2 | 0 | 0 | 1 | 1 | T1 | 6 | 5 | 9 | 8 | T1 | 5 | 12 | 9 | 9 |
| 3 | T1 | 0 | 0 | 1 | 1 | T2 | 3 | 4 | 8 | 8 | T4 | 3 | 3 | 10 | 3 |
| 4 | T3 | 0 | 0 | 1 | 0 | T3 | 0 | 0 | 7 | 5 | T2 | 0 | 0 | 10 | 3 |

期末の大問の分析

10%点: 2010年はT3の分割表, 2011年と2012年は相関と回帰が不要
 50%点: T3の分割表, T1の計算, T4のJMPが不要
 90%点: T1の計算が不要

| p | Var | MNM | Logi | LDF | QDF | Var | MNM | Logi | LDF | QDF | Var | MNM | Logi | LDF | QDF |
|---|-----|-----|------|-----|-----|-----|-----|------|-----|-----|-----|-----|------|-----|-----|
| 1 | T1 | 10 | 27 | 13 | 13 | T2 | 17 | 17 | 19 | 19 | T3 | 10 | 19 | 10 | 14 |
| 2 | T2 | 5 | 7 | 8 | 10 | T4 | 12 | 13 | 13 | 15 | T2 | 3 | 9 | 6 | 6 |
| 3 | T4 | 4 | 8 | 6 | 6 | T1 | 5 | 6 | 8 | 9 | T4 | 2 | 4 | 4 | 4 |
| 4 | T3 | 0 | 0 | 5 | 2 | T3 | 0 | 0 | 4 | 5 | T1 | 0 | 0 | 4 | 13 |
| 1 | T1 | 8 | 28 | 22 | 22 | T2 | 17 | 17 | 17 | 17 | T4 | 6 | 22 | 6 | 6 |
| 2 | T4 | 4 | 7 | 7 | 12 | T3 | 11 | 12 | 16 | 12 | T2 | 3 | 5 | 5 | 7 |
| 3 | T3 | 2 | 5 | 15 | 8 | T4 | 4 | 5 | 9 | 8 | T3 | 1 | 1 | 5 | 3 |
| 4 | T2 | 0 | 0 | 16 | 4 | T1 | 0 | 0 | 4 | 5 | T1 | 0 | 0 | 5 | 12 |
| 1 | T1 | 6 | 13 | 8 | 10 | T2 | 19 | 19 | 19 | 19 | T2 | 7 | 20 | 8 | 7 |
| 2 | T4 | 3 | 10 | 7 | 7 | T3 | 9 | 10 | 15 | 15 | T3 | 4 | 6 | 7 | 6 |
| 3 | T3 | 2 | 7 | 9 | 6 | T1 | 3 | 4 | 10 | 10 | T4 | 2 | 6 | 3 | 4 |
| 4 | T2 | 0 | 0 | 9 | 3 | T4 | 0 | 0 | 3 | 3 | T1 | 0 | 0 | 4 | 1 |

4.2 小問100問の分析

| | 年度 | P | MNM | Logi | LDF | QD | P | MNM | Logi | LDF | QD | P | MNM | Logi | LDF | QD |
|----|------|-----|-----|------|-----|-----|-----|-----|------|-----|----|-----|-----|------|-----|----|
| 中間 | 2010 | 6 | 0 | 0 | 2 | 1 | 12 | 0 | 0 | 2 | 4 | 13 | 0 | 1 | 4 | 13 |
| | | 96 | 0 | 0 | 0 | 109 | 96 | 0 | 0 | 0 | 61 | 96 | 0 | 0 | 0 | 13 |
| | 2011 | 12 | 0 | 0 | 2 | 107 | 15 | 0 | 0 | 3 | 6 | 9 | 0 | 0 | 6 | 9 |
| | | 98 | 0 | 0 | 0 | 107 | 98 | 0 | 0 | 0 | 61 | 98 | 0 | 0 | 0 | 9 |
| | 2012 | 6 | 0 | 0 | 7 | 114 | 19 | 0 | 0 | 0 | 3 | 15 | 0 | 0 | 0 | 12 |
| | | 100 | 0 | 0 | 0 | 114 | 100 | 0 | 0 | 0 | 67 | 100 | 0 | 0 | 0 | 12 |
| 期末 | 2010 | 12 | 0 | 0 | 5 | 111 | 12 | 0 | 1 | 4 | 4 | 11 | 0 | 1 | 6 | 13 |
| | | 99 | 0 | 0 | 0 | 111 | 99 | 0 | 0 | 0 | 62 | 99 | 0 | 0 | 0 | 13 |
| | 2011 | 8 | 0 | 0 | 4 | 4 | 13 | 0 | 0 | 6 | 7 | 8 | 0 | 0 | 2 | 12 |
| | | 97 | 0 | 0 | 0 | 110 | 97 | 0 | 0 | 0 | 62 | 97 | 0 | 0 | 0 | 12 |
| | 2012 | 10 | 0 | 0 | 1 | 115 | 10 | 0 | 0 | 5 | 4 | 9 | 0 | 0 | 6 | 12 |
| | | 97 | 0 | 0 | 0 | 115 | 97 | 0 | 0 | 0 | 63 | 97 | 0 | 0 | 0 | 12 |

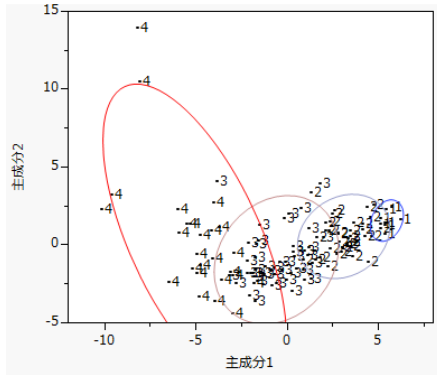
4.3 変数選択

変数選択法は, 線形分離可能なデータで問題がある.
 10%点: 24個中19個がより大きい次元を選ぶ
 50%点: 24個中17個がより大きい次元を選ぶ
 90%点: 24個中12個がより大きい次元を選ぶ

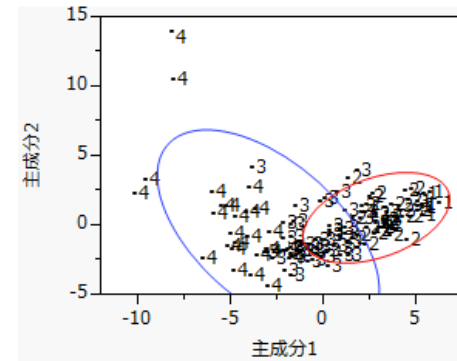
| | | | 10% | | | | | 50% | | | | | 90% | | |
|---------|----|----|-----|-----|-----|----|----|-----|-----|-----|----|----|-----|-----|-----|
| | F | Cp | AIC | BIC | MNM | F | Cp | AIC | BIC | MNM | F | Cp | AIC | BIC | MNM |
| 2010 中間 | 30 | 4 | 22 | 11 | 6 | 52 | 25 | 25 | 8 | 12 | 28 | 5 | 15 | 8 | 13 |
| 2010 期末 | 48 | 29 | 26 | 17 | 12 | 28 | 5 | 19 | 14 | 12 | 22 | 5 | 15 | 8 | 11 |
| 2011 中間 | 42 | 10 | 19 | 10 | 12 | 32 | 4 | 21 | 10 | 15 | 19 | 2 | 13 | 5 | 9 |
| 2011 期末 | 38 | 8 | 23 | 14 | 8 | 48 | 21 | 28 | 15 | 13 | 22 | 1 | 15 | 7 | 8 |
| 2012 中間 | 43 | 25 | 30 | 15 | 6 | 40 | 14 | 22 | 15 | 19 | 46 | 9 | 15 | 8 | 15 |
| 2012 期末 | 64 | 11 | 20 | 8 | 9 | 35 | 7 | 22 | 13 | 10 | 45 | 2 | 12 | 8 | 9 |

4.4 QDFが合格群を不合格と誤判別する現象

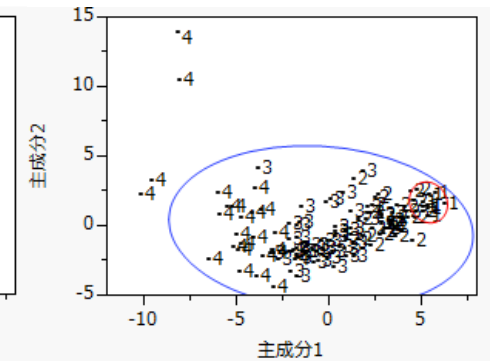
- 小問100問を主成分分析し、第1主成分と第2主成分で、スコアプロットを描く。
- 90%点, 50%点, 10%点で学生を第1群から第4群に分ける。
- なぜ第4群の成績の悪い学生の分散が大きいのか?
- なぜ, 90%点の合格群が10%点の不合格群に誤判別されるのか?



50%点の合否判定



90%点の合否判定



4.5 QDFが合格群の全てを誤判別する理由と判別理論の修正

- LDFやQDFは、計算式の中に分散共分散行列の逆行列を含む
- QDFと正規化法: 分散共分散行列の対角要素を修正することで、ダーティな判別に対応
- 単に一定値をとる変数に $N(0, 10^{-6})$ を加えれば良い。

| | LDF | QDF | 平均の差の検定 |
|-----------------------|-----|--------------|---------|
| 2群の値が同じ | 省く | 省く | 省く |
| 2群が別の一定値(判別に重要) | 省く | 省く | 省く |
| 一方が同じで、他方がばらつく(判別に重要) | 計算 | 合格群を不合格群に誤判別 | 計算 |
| SPSS | 省く | 省く | 省く |

| p | VAR | MMM | pLDF | pQDF | 修正 | VAR | MMM | pLDF | pQDF |
|----|------|-----|------|------|----|------|-----|------|------|
| 1 | x85 | 10 | 14 | 14 | 14 | x92 | 12 | 12 | 12 |
| 2 | x15 | 6 | 14 | 114 | 28 | x42 | 8 | 8 | 12 |
| 3 | x68 | 5 | 8 | 114 | 28 | x21 | 5 | 5 | 12 |
| 4 | x47 | 3 | 8 | 114 | 28 | x54 | 4 | 8 | 12 |
| 5 | x7 | 1 | 4 | 114 | 9 | x65 | 1 | 3 | 12 |
| 6 | x32 | 0 | 5 | 114 | 3 | x100 | 1 | 3 | 12 |
| 7 | x20 | 0 | 3 | 114 | 0 | x83 | 1 | 3 | 12 |
| 14 | x98 | | 3 | 114 | | x1 | 1 | 1 | 12 |
| 15 | x5 | | 1 | 114 | | x62 | 0 | 1 | 12 |
| 16 | x1 | | 0 | 114 | | x3 | | 1 | 12 |
| 18 | x38 | | | 114 | | x60 | | 0 | 12 |
| 19 | x6 | | | 114 | | x96 | | | 12 |
| 20 | x89 | | | 114 | | x22 | | | 12 |
| 21 | x100 | | | 114 | | x40 | | | 12 |

2013年の中間の10%と90%

- 一定値をとる設問に $N(0, 10^{-9})$ を加えるだけで解決
- しかし、分散共分散行列の(対角要素)の修正という研究スタイルを変えず
- データがばらつかないことを認めるべき
- 90%点は, X92($t=16.0, 12/34$), X65($t=12.2, 12/48$), X83($t=7.85, 12/72$)

5. 終わりに

- 判別分析に関する多くの問題は、最適線形判別関数で全て解決[12].
- 試験の合否判定データ
 - 良質なMNM=0の研究データ
 - 大問による合否判定で、MNM=0になる設問と不要な設問で、試験の質や学生の理解度が分析.
 - 統計入門のような入門科目の簡単な設問の場合、100問中6問ぐらいで合否判定可.
 - 入試データの統計分析をなぜ行わないのか？
- 大学教育に、統計家は積極的にかかわろう.

文献

- [1] Fisher, R.A. (1936). The Use of Multiple Measurements in Taxonomic Problems, *Annals of Eugenics*, 7, 179–188.
- [2] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27–38.
- [3] Flury, B. & Rieduy, H. (1988). *Multi-variate Statistics: A Practical Approach*, Cambridge University Press, Cambridge.
- [4] Shinmura, S. (2000). A new algorithm of the linear discriminant function using integer programming, *New Trends in Probability and Statistics*, 5, 133–142.
- [5] Shinmura, S. (2011). Beyond Fisher's Linear Discriminant Analysis - New World of Discriminant Analysis -, *ISI2011 Proceedings*, 1–6.
- [6] 新村秀一, 三宅章彦 (1983). 重回帰分析と判別解析のモデル決定 (1) - 19変数をもつCPDデータの多重共線性の解消-, *医療情報学*, 3/3, 507–124.
- [7] 新村秀一 (1998). 数理計画法を用いた最適線形判別関数, *計算機統計学*, 11/2, 89–101.
- [8] 新村秀一 (2004). JMP活用 統計学とっておき勉強法. 講談社, 東京.
- [9] 新村秀一 (2007). JMPによる統計レポート作成法. 丸善.
- [10] 新村秀一 (2007). 数理計画法による判別分析の10年, *計算機統計学*, 20(1/2) 53–94.
- [11] 新村秀一 (2007). ExcelとLINGOで学ぶ数理計画法. 丸善.
- [12] 新村秀一 (2010). 最適線形判別関数. 日科技連出版社.
- [13] 新村秀一 (2011). 問題解決学としての統計入門. 第7回統計教育の方法論ワークショップ-問題解決力育成を目指した統計教育の方法論-, 1–10.
- [14] 新村秀一 (2011). 合否判定データにおける判別分析の問題点. *応用統計学*, 3, 157–173.
- [15] 新村秀一 (2011). 数理計画法による問題解決法. 日科技連出版社.
- [16] 新村秀一 (2012). Fisherの判別分析を超えて. 2012年SASユーザー会論文集, 349–361.
- [17] 新村秀一 (1984). 医療データ解析, モデル主義, そしてOR. *オペレーションズ・リサーチ*, 29-7, 415–421.
- [18] Miyake, A. & Shinmura, S. (1976). Error rate of linear discriminant function, *F.T. de Dombal & F.Gremy*, editors 435–445, North-Holland Publishing Company.
- [19] 田口玄一 (1999). タグチメソッドわが発想法. 経済界, 東京.
- [20] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.
- [21] 新村秀一, ユンイエブ (2007). OLDFとSVMの比較研究(4) — 種々のデータによるSVMとの比較 —, *成蹊大学経済学部論集*, 37-2, 89–119.
- [22] 新村秀一・鈴木隆一郎・中西克己 (1983). 各種判別手法を用いた医療データ解析の標準化 — マンモグラフィによる乳癌の診断 —. *医療情報学*, 3-2, 38–50.