

2010年と2011年度の統計入門の分析 — 2011年度の半舷授業の影響分析から センター試験の分析へ—

新村秀一
成蹊大学 経済学部

1. はじめに

- 2010年度から、成蹊大学経済学部では基本統計量と相関・単回帰分析を中心にした「統計入門」を1年次生全員(500人)に対し必修。
 - 4人の教員が担当し、筆者も125名の1年次生と若干名の2年次生以上を教える。
 - 中間試験と期末試験は、10択100問のマークセンス試験で行う。
 - 合否判定は、4人の教員で受験者の得点分布の10%点で足切りすることに決めた。
- 統計の研究者が、自分の試験を統計分析しFD活動に役立てないのはおかしい。
 - そこで、2010年度の中間試験の結果を、10%点と判別超平面の近辺に多くの受験生がくる50%点を合否判定の基準として、FD(Faculty Development)を兼ねて種々の観点から分析。
 - 合否10%点(合格最低点は48点)を目的変数とし、100問の小問で判別を行うと、わずか6問(6点)で合否判定できた(新村(2011c))。
- 合否判定は、最適線形判別関数の応用研究に最適なテーマ
 - 試験のデータを判別分析しても、得られた判別関数を一般的に試験内容の異なる次の試験に適用できない。
 - しかし、大問で合否判定可能な最小の設問数、あるいは小問で合否判定に選ばれた設問が、試験問題の質を説明できることが期待できる。
 - そこでこの点を大学入試センターに2009年に提案し、2010年度に都内の国立大学1年生で研究用に集めた105種類の本試験と追試験105問題の提供を受けた。
 - 昨年末、応用統計学会主催の「学力試験データの統計解析」の特集号とセミナーで2010年度の分析結果を報告。
 - 昨年中に、すべてのセンター試験の分析を終了。
 - 今回は自分の試験で、試験データの質判定の分析方法を検討して、今回報告。

分析方法

- センター試験データを分析する前に、試験内容を熟知している「統計入門」で、従来の判別手法(LDF, 2次判別関数, 名義ロジスティック回帰)の合否判定の問題点の検証と試験問題の質に関して研究した.
 - 新村(2011a)では, 2010年度の間試験と期末試験のデータを10%点と50%点で合否判定を行った.
 - 本研究では10%点と対称な90%点の合否判定を加える.
 - 合否90%点は, 難解な資格試験を想定.
 - また, 2011年度の試験結果が得られたので, これも付け加え
 - 2011年度は計画停電の影響もあり, 15回の授業日数が途中から11回の半腕授業に変更になり, この影響も分析する.
 - また, 今後105種類のセンター試験データを統一的に解析するための分析方法の検討を行った.
 - ただし, 2012年1月に外部発表を行わない依頼を受け, 投稿中の日本分類学会誌の論文2編, 日独分類シンポジュームのabstractを取り下げた.

2. 統計入門の概略

- 授業の目的: 統計の入門科目として基本統計量と相関と単回帰と分割表を教えること.
 - テキスト:『JMP活用 統計学とっておき勉強法』
 - 第1部では $(x,y)=(0,1),(1,1),(1,3),(2,3)$ という2変数*4件の簡単なデータで, 基本統計量, 相関と単回帰, 分割表の計算を教え, 統計量の意味を説明.
 - 第2部では「学生の成績データ」を用いて, JMP(Sall et al.(2004), 新村(2007))の出力結果を用いて, 統計量の具体的な意味を説明.
 - 用いているデータは, 40人の学生の成績, 勉強時間などの7変数.

表1 2010年と2011年度の講義内容の比較

週	2010年	2011年
1	PowerPointで概論	PowerPointで概論
2	最頻値, 中央値, 平均値	最頻値, 中央値, 平均値
3	範囲, 四分位範囲,SD,CV	範囲, 四分位範囲,SD,CV
4	学生データの解釈	学生データの解釈
5	正規分布	正規分布と推測統計
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	Excelで相関の計算
9	Excelで相関の計算	単回帰分析
10	単回帰分析	分割表と独立性の検定
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	補講	
15	期末試験	

授業の内容

- 分布の代表値, 分布のバラツキ, 分布の形, 変動係数を, 4件のデータで手計算.
- 散布図と相関係数, 単回帰式と分散分析表, 分割表の計算.
- 推測統計
 - 平均で標準誤差と95%信頼区間の説明.
 - 正規の当てはめの適合度検定
 - Fisherの直接確率で, 推測統計の仕組み.
 - 相関係数のp値.
 - 分散分析表のp値.
 - 統計ソフトの出力の解釈.

表2 両年度の試験結果

次元は合否判定できた設問数(得点)。2010年の中間の10%点は、合格最低点が48点で、6問で合否判定可能。期末は試験内容が豊富なので、中間試験より最大で8点悪い。ただし、合否水準を上げると成績の優秀な学生は影響を受けない。2011年度は、試験内容を5点ほど難しくしているが、期末は授業回数が8回が5回に減ったのに、成績は少し良くなっている。

		2010年度		2011年度	
		点	次元(比=次元/合格点)	点	次元(比)
中 間	10%点	48	6 (0.13)	42	12 (0.29)
	50%点	66	12 (0.18)	61	15 (0.25)
	90%点	82	13 (0.16)	79	9 (0.11)
	最高点	93		88	
期 末	10%点	40	12 (0.30)	43	8 (0.19)
	50%点	60	12 (0.20)	60	13 (0.22)
	90%点	82	11 (0.13)	81	8 (0.10)
	最高点	91		99	

なぜ、2011年度の期末の成績が良いか？

99点の女子学生(中間試験は70点台)を調査

- 1) 期末試験の範囲の授業になってから、私が授業の最初に毎回予習してきたかと聞くので、テキストや配布資料で予習。
- 2) 中間試験の後の授業から、黒板の板書はやめてアップロードするようになった講義ノートで、最初はさっと眺めただけで効果はなかった。2回目からは印刷しノート代わりにしたので中間試験の範囲(特に正規分布表と偏差値の関係)より理解が容易だった。

結局、半舷授業の悪影響を克服して好成績であった理由。

- 1) 講義ノート等の資料を事前に開示するという工夫と、
- 2) 多くの学生がテキストやPowerPointによる事前配布資料の予習、
- 3) Excelの計算式の開示が良かった。

さらに、予習が一番大学で初めて習う科目に最適ということを毎回いったことが良かった。

試験のデータを統計分析しないと、ここまでの核心には至らなかった。

3. 大問の分析

(大問を内容で組み直した. 中間はT2が易しく, 期末のT1は成績不振者にとっては難しい. 中間のT2は90%点の合否判定に不要なことが分かる. 試験問題の質評価に利用可能)

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

表4 2010年度(上)と2011年度(下)の中間の大問の合否判定(MNMからLogiは誤分類数)

p: 変数増加法の説明変数の数, 両年度の90%点では, MNMが3変数で合否判定可能. すなわち, T2が成績優秀な判別に不要. 10%点では, 成績不振者にとって正規分布と偏差値の関係が難しい.

P.	Var.	MNM	LDF.	QD.	Logi.	Var.	MNM	LDF.	QD.	Logi.	Var.	MNM	LDF.	QD.	Logi.
1.	T4.	6.	11.	11.	9.	T4.	16.	16.	16.	16.	T3.	10.	24.	24.	27.
2.	T2.	2.	11.	9.	6.	T3.	9.	12.	12.	10.	T4.	5.	20.	11.	10.
3.	T1.	1.	8.	5.	3.	T1.	2.	5.	6.	2.	T1.	0.	20.	10.	0.
4.	T3.	0.	9.	2.	0.	T2.	0.	3.	6.	0.	T2.	0.	20.	11.	0.

P.	Var.	MNM	LDF.	QD.	Logi.	Var.	MNM	LDF.	QD.	Logi.	Var.	MNM	LDF.	QD.	Logi.
1.	T2.	9.	15.	15.	17.	T4.	9.	9.	9.	9.	T3.	6.	14.	14.	7.
2.	T4.	4.	11.	9.	9.	T1.	4.	5.	7.	4.	T4.	1.	14.	6.	1.
3.	T1.	0.	9.	10.	0.	T3.	1.	3.	3.	2.	T1.	0.	13.	5.	0.
4.	T3.	0.	9.	11.	0.	T2.	0.	3.	3.	0.	T2.	0.	14.	9.	0.

表4 期末の大問の合否判定

LDFとQD(2次判別)は, 4試験で全問を利用しても合否判定できない. 計算問題は, 中間と同じく成績優秀者にとって楽勝のため10%、50%、90%で後で取り込まれる. 2010年度と2011年度で, T4の傾向が違うのは, 2011年度の期末にテキスト購入が増えたためと考えている.

P _o	Var _o	MNM _o	LDF _o	QD _o	Logi _o	Var _o	MNM _o	LDF _o	QD _o	Logi _o	Var _o	MNM _o	LDF _o	QD _o	Logi _o
1 _o	T1 _o	10 _o	21 _o	21 _o	27 _o	T2 _o	17 _o	19 _o	19 _o	17 _o	T3 _o	10 _o	22 _o	19 _o	19 _o
2 _o	T2 _o	5 _o	16 _o	17 _o	7 _o	T4 _o	12 _o	13 _o	15 _o	13 _o	T2 _o	3 _o	17 _o	13 _o	9 _o
3 _o	T4 _o	4 _o	14 _o	11 _o	8 _o	T1 _o	5 _o	8 _o	9 _o	6 _o	T4 _o	2 _o	16 _o	9 _o	4 _o
4 _o	T3 _o	0 _o	16 _o	5 _o	0 _o	T3 _o	0 _o	5 _o	5 _o	0 _o	T1 _o	0 _o	14 _o	13 _o	0 _o

P _o	Var _o	MNM _o	LDF _o	QD _o	Logi _o	Var _o	MNM _o	LDF _o	QD _o	Logi _o	Var _o	MNM _o	LDF _o	QD _o	Logi _o
1 _o	T1 _o	8 _o	22 _o	22 _o	28 _o	T2 _o	17 _o	17 _o	17 _o	17 _o	T4 _o	6 _o	17 _o	17 _o	22 _o
2 _o	T4 _o	4 _o	7 _o	12 _o	7 _o	T3 _o	11 _o	16 _o	12 _o	12 _o	T2 _o	3 _o	11 _o	7 _o	5 _o
3 _o	T3 _o	2 _o	15 _o	8 _o	5 _o	T4 _o	4 _o	9 _o	9 _o	5 _o	T3 _o	1 _o	11 _o	9 _o	1 _o
4 _o	T2 _o	0 _o	16 _o	4 _o	0 _o	T1 _o	0 _o	6 _o	4 _o	0 _o	T1 _o	0 _o	11 _o	12 _o	0 _o

表5 一元配置の分散分析

	2010 中間	2010 期末	2011 中間	2011 期末
総合 得点	-	-	-	-
T1	-	AABC	-	-
T2	AAAB	-	AAAB	AAAB
T3	ABCC	-	ABCC	ABCC
T4	-	ABCC	-	-

- 総合得点と4個の大問の得点を, 90%, 50%, 10%点で受講生を4水準に分け, 一元配置の分散分析で4群の平均値の差を検定.
- 「-」は, 4カテゴリーの平均値すべてに差がある. 設問として切れ味が良い. 両年度の中間では, 基礎統計量の概念と解釈で平均に差. 両年度の期末が異なるには, 2011年度はテキストの予習効果の影響と考える.
- 平均値の大きな水準から差があった場合に, A, B, C, Dのアルファベットで表示.
- 水準間で差がない試験は, 受験生の能力を適切に評価していない.
- 2010年と2011年の中間試験で, T2はAAAB,
 - 10%未満と10%以上の平均だけに差がある.
- 一方, 2011年の期末試験のT1とT4は, ABCDと4水準で平均に差があった.
 - しかし, 2010年度ではAABCとABCCであり, 3水準で平均に差がある.

4. 小問100問の分析

4.1 誤分類数と線形分離可能な次元

- ・10%点は6問, 50%点は12問, 90%点は13問で合否判定可能. しかし, LDFとQDは合否判定できない.
- ・QDFは, いずれも一方の群を全て誤判別. 10%点では合格の9割が不合格の1割に誤判別. 理数科目の特徴か?
- ・合否判定できた設問数のパターンが試験問題の質を?

表6 2010年度の間試験の判別結果(次元は右下がり傾向)

P.	Var.	Type	MNM.	LDF.	QD.	Logi.	Var.	Type	MNM.	LDF.	QD.	Logi.	Var.	Type	MNM.	LDF.	QD.	Logi.
1.	36.	Tfn.	11.	11.	11.	11.	93.	tFn.	28.	28.	28.	28.	57.	tfN.	13.	19.	19.	18.
5.	17.	.	2.	7.	9.	3.	97.	tFn.	10.	10.	13.	10.	55.	tfN.	6.	9.	8.	8.
6.	61.	Tfn.	0.	2.	1.	0.	99.	tFn.	9.	9.	12.	9.	59.	tfN.	5.	8.	7.	5.
11.	3.	38.	tFn.	1.	4.	5.	1.	1.	tfN.	1.	6.	13.	2.
12.	65.	22.	tFn.	0.	2.	4.	0.	3.	tfN.	1.	6.	13.	1.
13.	72.	67.	63.	tfN.	0.	4.	13.	1.
14.	99.	34.	8.	.	.	5.	13.	0.
96.	46.	.	.	0.	109.	0.	44.	.	.	0.	61.	0.	30.	.	.	0.	13.	0.

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. 左から右は, 10%, 50%, 90%点で受験生を1:4:4:1に4分割した2次元95%正規確率楕円を当てはめた. 左の成績不振な10点未満の学生は, 正解率が低い, 正解のパターンが大きいので分散共分散が大きい. 成績が良くなるにつれ, 合格最低点が上がるが, 分散共分散は小さくなる. この場合, 10%以上の受験生が, 10%以下の群に誤判別される. 減少として解明できたが, 誰か理論的に解明してください. 鏡餅のような4段重ねの場合, 一方の群に全て誤判別されることは起きない.

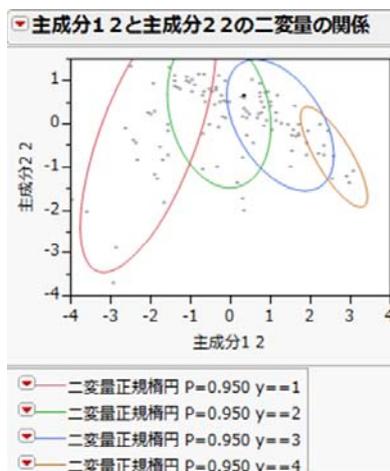


表7 2010年度の期末試験の判別結果

- 1個の設問が全員正解で、99変数まで求まった。
- 10%点と50%点では12変数、90%点では11変数で合否判定可能(Flatな傾向)
- 従来の判別関数は、判別境界上のケースの問題を正しく扱ってこなかった。
 - 50%点では名義ロジスティック回帰が31変数まで誤分類数が1。これは、全ての判別関数が判別境界上のケースの判別を正しく処理していないため。
 - 90%点では、MNMが合否判定できる11変数で誤分類数が0にならないで、12変数で0。これも利用者が、陽性に指定した方に判別境界のケースを無条件に判別するためにおきる。
- LDFと2次判別関数は、MNM=0のモデルで誤分類数は0でない。
- また99変数では、QDFはいずれも一方の群を誤判別している。10%点では、11変数になると突然、9割を不合格群に誤判別する。90%点では、合格群の1割の分散共分散が小さいので、9割の不合格群に誤判別。

表7 2010年度の期末試験の判別結果 (次元は平坦傾向)

P.	Var.	Type	MNM	LDF	QD	Logi	Var.	Type	MNM	LDF	QD	Logi	Var.	Type	MNM	LDF	QD	Logi
1.	18.	Tfn.	12.	15.	15.	15.	22.	tFn.	12.	26.	26.	26.	68.	tFn.	12.	29.	29.	29.
11.	1.	Tfn.	1.	5.	111.	1.	90.	tFn.	1.	6.	9.	1.	34.	tFn.	0.	6.	13.	1.
12.	2.	Tfn.	0.	5.	111.	0.	19.	tFn.	0.	4.	4.	1.	70.	.	.	4.	13.	0.
99.	73.	.	.	0.	111.	0.	71.	.	.	0.	62.	0.	58.	.	.	0.	13.	0.

表8 2011年度の中間試験の判別結果

- 2個の設問が全員正解で98変数まで求まった。
- 3水準で12変数、15変数、9変数で合否判定可能(「V傾向」)
- このような傾向を比較することで、試験の質が分かると考えている。
- TypeのTは10%点の合否判定に選ばれた場合、選ばれない場合はtで表記。F/tとN/nは、50%点と90%点を表す。10%点の13問目の設問はTfnであり、10%点の合否判定にしか役に立たない、計算問題である。

P.	Var.	Type	MNM	LDF	QD	Logi	Var.	Type	MNM	LDF	QD	Logi	Var.	Type	MNM	LDF	QD	Logi
1.	13.	Tfn.	9.	9.	9.	9.	84.	tFn.	19.	19.	19.	19.	54.	tFn.	9.	28.	9.	28.
9.	22.	Tfn.	2.	2.	107.	7.	83.	tFn.	2.	5.	5.	2.	94.	tFn.	0.	6.	9.	0.
10.	17.	Tfn.	2.	3.	107.	2.	23.	tFn.	2.	5.	5.	2.	56.
11.	90.	Tfn.	1.	4.	107.	2.	82.	tFn.	2.	5.	5.	3.	52.
12.	14.	Tfn.	0.	2.	107.	0.	63.	tFn.	1.	5.	5.	1.	82.
13.	18.	26.	tFn.	1.	5.	6.	1.	40.
14.	78.	52.	tFn.	1.	4.	5.	1.	55.
15.	69.	98.	.	0.	3.	6.	0.	78.
98.	87.	.	.	0.	107.	0.	91.	.	.	0.	61.	0.	51.	.	.	0.	9.	0.

2011年度の期末試験

- 3水準で8変数, 13変数, 8変数で合否判定可能
- 次元は90%点の次元が小さくなる「V傾向」.
- LDFと2次判別関数は, $MNM=0$ のモデルで誤分類数は0でない.
- また, LDFの97変数の誤分類数は0である.
- 2次判別関数の誤分類数は, 3水準で110, 62, 12個であり, いずれも一方の群を誤判別.

4.2 変数選択(2010年度中間)

- 表9は, 変数増加法(F), Cp統計量, AIC, BICで選ばれた説明変数の個数
- MNMは最適線形判別関数で最初に $MNM=0$ になるモデルの次元数。「比」は, 次元数を合格点で割ったものである.
- 4試験の各水準で16個ある変数選択法のうち, 下線を引いた12個, 12個, 8個がMNMMで合否判定できる次元より大きなモデルを選ぶ傾向
- 少なくとも, $MNM=0$ という非常にはっきりした意味をもつモデルを選ぶのに適した逐次変数選択の検定法はない

表9 変数増加法, Cp統計量, AIC, BICとMNMで選ばれた説明変数の個数。

	10%						50%						90%						傾向
	F	Cp	AIC	BIC	MNM	比	F	Cp	AIC	BIC	MNM	比	F	Cp	AIC	BIC	MNM	比	
2010 中間	30	4	22	11	6	0.13	52	25	25	8	12	0.18	28	5	15	8	13	0.16	右下がり
2010 期末	48	29	26	17	12	0.30	28	5	19	14	12	0.20	22	5	15	8	11	0.13	平坦
2011 中間	42	10	19	10	12	0.29	32	4	21	10	15	0.25	19	2	13	5	9	0.11	V
2011 期末	38	8	23	14	8	0.19	48	21	28	15	13	0.22	22	1	15	7	8	0.10	V

4.3 設問のType分類

- 合否判定に選ばれた変数と選ばれなかった変数を、10%点ではTとtで、50%点ではFとfで、90%点ではNとnで表わす。
- 全部のTypeは、表に示す通り8通りある。
- 表の最上段の4個のTypeは90%点で合否判定に関係した設問のTypeを表す。
 - 「tfn」が一番多い。どの水準の合否判定に利用できる設問は、TFNが0であり、理にかなっている。
- 中段の2個は50%点で選ばれ、「tFn」が50個と一番多い。
- 10%でしか選ばれない「Tfn」は36問。
- 400問中272問が合否判定に関係しない。これらは不要というわけではない。

表 10 合否判定に選ばれた変数の Type 分類

Type	2010 中間	2010 期末	2011 中間	2011 期末	計
TFN	0	0	0	0	0
TfN	1	0	0	0	1
tFN	0	0	1	0	1
tFn	12	11	8	8	39
TFn	0	0	0	1	1
tFn	12	12	14	12	50
Tfn	5	12	12	7	36
tfn	70	65	65	72	272

表11 正答率のTypeによる一元配置(2011)

- 表11は、2011年度の中間試験(上)と期末試験(下)の小問の正答率を、Typeで一元配置の分散分析、
 - 平均に差があるか否かを検定
 - 大問ではこの分析は有効と考える。しかし1問1点の本試験では、有効でないが、設問の得点が高い問題では有効かもしれない。
- 中間試験では、(Tfn, tfn)とtfnの正答率の平均に差がある。
- 期末試験では、(Tfn, tFn, tfn)とtfnに差がある。
- 共通して言えることは、Tfn, tFn, tfnの順に平均が小さくなる。
- 四分位範囲をみると、いずれもかなり大きいように考える。

表 11 正答率の Type による一元配置

水準(検定)	平均	MIN	10%	25%	50%	75%	90%	MAX	Q3-Q1
Tfn (A)	76.0	40.0	40.0	70.0	79.0	87.0	93.0	93.0	17.0
tfn (A)	63.2	10.0	29.0	46.0	66.0	85.5	93.0	99.0	39.5
tFn (AB)	49.0	5.0	5.9	35.0	57.5	64.5	74.8	76.0	29.5
tfn (B)	40.5	23.0	24.2	31.0	41.0	45.0	69.2	74.0	14.0

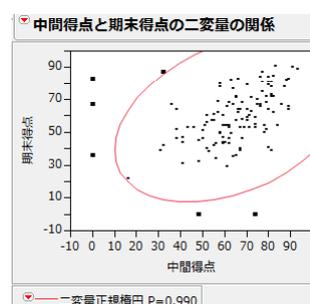
水準(検定)	平均	MIN	10%	25%	50%	75%	90%	MAX	Q3-Q1
Tfn (A)	77.3	61.0	61.0	63.5	77.5	90.8	93.0	93.0	27.3
tFn (A)	61.8	43.0	46.0	53.5	61.0	66.8	85.1	92.0	13.3
tfn (A)	61.7	12.0	27.9	45.5	60.0	79.0	93.0	100.0	33.5
tfn (B)	36.6	24.0	24.0	25.3	32.5	49.5	57.0	57.0	24.3

5. まとめ

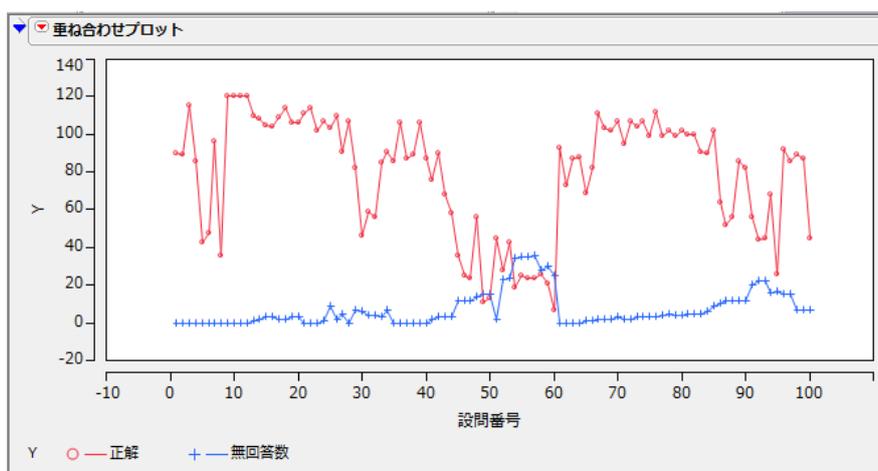
5.1 試験結果の評価

- 2010年度と2011年度の中間試験と期末試験の比較を行った.
- 2011年度は、中間試験の授業が7回から6回、期末試験が8回から5回に半減授業.
- 2011年度の中間と期末は、5点ほど難しくしたが、期末試験は2010年度の期末、2011年度の中間に比べて成績が上がった.
- これは、次の点による効果と考える.
- 授業の前に、予習の重要さを毎回伝え、教科書の購入者が3割ほどに増え、成績上位者が予習をしてきた.
- ノートを取っていたが、間違いと集中力に欠くので、板書する内容を事前にUPLし、ノート代わりにすることを推奨.
- Excelのセルに定義した数式をUPLした.

下図は2010年度の得点分布. 中間が30点で期末が80点以上の学生に注目. 1年次の授業では、FDを兼ねて試験結果を分析し、このような学生を発見し、指導することが重要. また中間/期末が欠席で70点以上の学生もいる. 統計家は試験結果を統計分析し、学生に示すことが重要.



2010年度の中間試験の正解と無回答数の比較 この簡単な分析結果で出題内容の反省ができる



5.3 最適線形判別関数

- 新村は、誤分類数最小化基準による最適線形判別関数を開発し、判別分析の全ての問題を解決した(最適線形判別関数,日科技連出版社, 2010) .
- 改定 IP-OLDFの定義

$$MIN = \sum e_i ;$$

$$y_i * (x_i' b + b_0) \geq 1 - c * e_i ; \quad (b, b_0) \text{は, 判別係数. } e_i \text{はケース } x_i \text{に対応した0/1の整数変数. } c \text{を大きな整数変数 (例えば10000) にすると, ケースが誤分類されなければ 「 } y_i * (x_i' b + b_0) = 1 ; \text{」, 誤分類されれば 「 } y_i * (x_i' b + b_0) = -9999 ; \text{」 という判別超平面の2者択一の選択になる. 誤分類されるケースが-9999に引っ張られ 「 } y_i * (x_i' b + b_0) = 0 ; \text{」 すなわち超平面上にケースは来なくなる. 「 } MIN = \sum e_i ; \text{」 で誤分類数を最小化する.}$$
- 判別係数と誤分類数の関係が、初めて分かった.
- 判別超平面上にケースが来ないようにできるのは、改定IP-OLDFだけで、SVMではできない.
- MNMの単調減少性: ある説明変数でMNM=0になれば、それを含むすべてのモデルでMNM=0になる. すなわち、次の単純な関係があり、これが試験の合否判定の応用研究につながった. $MNM_p \geq MNM_{(p+1)}$

5.4 最適線形判別関数の懸念

- 内部標本(教師データ)のケース数を最小化しているので、外部標本(評価データ)に適用すると、誤分類確率が悪くなるという懸念.
- しかし、次に示す4種の実データで100組のリサンプリングデータを作成し、100重交差検証法で検証し、LDFとロジスティック回帰に比べ、評価データでも誤分類確率は少なかった.
- 4種類の実データ
 - Irisデータ(Fisherの仮説を満たす)
 - 銀行紙幣データ(2変数でMNM=0)
 - CPD(多重共線性)
 - 学生データ(一般位置にない)
- 100組のリサンプリング・データを生成. 100重交差検証法で、教師データと評価データで評価
 - 139組の異なった判別問題で、好成績
 - 判別係数と誤分類数の95%信頼区間
- リサンプリングデータに偏りがあるという指摘があるが、同じデータを使って、LDFとロジスティック回帰は悪いという真実は重要.
- この結果を得て、1997年から12年間行ってきた最適線形判別関数の研究を完成.
- 応用研究として、合否判定データは誤分類数が0の出生のはっきりしたデータで、他の研究者が容易に確認できる. ドイツの統計学者のFluryらは、スイス銀行1000フラン紙幣の真札・偽札のデータを集め判別分析の本を書いているが、線形分離可能なことを発見できなかった.

5.4 100重交差検証法

LDFとロジスティック回帰の平均誤分類確率から, MNMの平均誤分類確率を引いた差の最大/最小値

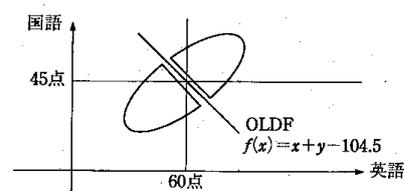
	LDF-IPLP				Logi-IPLP			
	教師		評価		教師		評価	
	最小値	最大値	最小値	最大値	最小値	最大値	最小値	最大値
アイリス(15)	0.55	5.23	-0.60	2.36	0.59	5.31	-0.84	1.85
銀行(63)	0.00	3.63	-0.01	4.35	-0.28	3.47	-0.04	4.43
学生(31)	1.46	8.61	-1.29	7.11	-2.12	6.48	-2.89	5.59
CPD(26)	3.05	7.28	2.21	6.15	0.13	3.43	0.29	1.74

5.5 応用研究(合否判定)

- 応用研究として, 従来の判別関数が線形分離可能なデータを認識できないことに焦点を絞った.
- 試験の合否判定は, 最良の研究テーマである. 誤分類数0の判別関数は自明.
- N個の設問の得点Xが50点以上で合格と判定すると,
 $X=x_1+\dots+x_n>49.5$ で合格.
- 一般的に線形分離可能なデータを探すには, 最適線形判別関数を用いて, 多くの人のボランティアが必要. これまで, 大学卒業後, 医療診断の分析を行ってこなかったが, その中には誤分類確率0のものがあった可能性.

合否判定に対する批判

- 受験生が仮に正規分布しているとして, それを合否に分ければ, 正規分布でないのでFisherのLDFを適用してはいけない
- LDFは線形分離可能なことを認識できないので, その判定ができない. また, 与えられたデータ全てに適用できる最適線形判別関数があるので, LDFは不要といっているのと同じ.
- 判別分析は, 外的基準が別の観点から決められ, それを他の計測値で判別する手法
- これは, 単にFisherの判別の延長での主張. 各種格づけは得点を積み上げ, それがある基準以上か否かで判定. これも従来の判別分析の規範で考えている.
- **合否判定は, 次の試験に応用できない.**
- **合否判定できる最小の設問数の情報で, 試験の質が評価できる.**



5.6 まとめ

- 1998年から12年間行ってきた最適線形判別関数の応用研究として、試験の合否判定における問題点を検討した。
- Fisherの判別分析の終焉
 - 判別する2群が正規分布と仮定し分散共分散から導かれるLDFと2次判別関数は、大問では4問全てを用いても合否判定ができなかった。
 - 小問では、 $MNM=0$ になる最小の説明変数のモデルで合否判定ができなかった。
 - フルモデルでは、LDFの誤分類数は0になったが、2次判別関数は一方の群が全て誤判別された。
- 誤分類確率が、LDFでは最大0.34、QDFでは0.94である衝撃。
 - 従来、このような誤分類確率が得られる場合、誰も線形分離ということは認識できない。
 - 医学での判別関数による研究を見直す必要はないか。
 - ゲノム判別で、少ないケース数で多くの説明変数をもつデータの分散共分散の推定の研究は、問題。
 - 正則化法は問題。
 - 若い研究者が、LDFの判別境界を動かすと誤分類確率が少なくなるという研究をしているが、やめるようアドバイスをするのが年配者の務め。

5.7 名義ロジスティック回帰

- 名義ロジスティック回帰は、2群が正規分布していることを仮定せず、教師データの分布に合わせて導き出される。
 - このため、多くの事例で誤分類確率が小さくなるため、LDFと2次判別関数に代わって医療や金融で近年多用されている。
 - また、回帰係数の推定値が全て不安定になるまで変数を追加していき、誤分類数が0になると、多くの場合に最適線形判別関数が見つけた最小次元の特徴空間を見つけた。
 - しかし、これまで判別係数の推定が不安定になればそれ以前で変数選択を止めるのが常識であった。
 - このため、統計的判別関数で唯一線形分離可能なデータをかろうじて認識できる名義ロジスティック回帰であるが、線形分離可能なデータを見逃してきた可能性が高い。

5.8 お願い

- 各大学の入試試験, 各教員の試験データで筆者の提案した方法で試験問題の質評価という新しい統計の研究分野を確立しませんか.
- 大学入試センターに変わる, 試験データの提供, あるいは共同研究
 - 予備校で, 私立3教科型のデータを都内私立大生のアルバイトで収集し, 研究を行う.
 - 技術士試験などの過去のデータを用いた, 試験の質を検証する.

参考文献

- Firth,D.(1993). Bias reduction of maximum likelihood estimates, *Biometrika*, **80**, 27-38.
- Sall, J.P., Creighton, L. and Lehman, A. (2004). JMPを用いた統計およびデータ分析入門(第3版), SAS Institute Japan (株).[新村秀一監修].
- Shinmura,S. (2011).Beyond Fisher's Linear Discriminant Analysis -New World of Discriminant Analysis -, ISI2011 Proceedings, 1-6.
- Shinmura,S.and Ishioka,T.(2012). Serious problems of Discriminant Analysis,第4回日独分類シンポジウム・アブストラクト.
- 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.
- 新村秀一(2007a). 数理計画法による判別分析の10年,計算機統計学, **20/1&2**, 59-94.
- 新村秀一(2004). JMP活用統計学とっておき勉強法,講談社.
- 新村秀一(2007d). JMPによる統計レポート作成法, 丸善.
- 新村秀一(2009b). 数学のできる問題解決法, 成蹊大学一般研究報告, **42/4**, 1-52.
- 新村秀一(2010). 最適線形判別関数, 日科技連出版社.
- 新村秀一(2011a). 合否判定データによる判別分析の問題点,応用統計学, **40/3**, 157-172.
- 新村秀一(2011b). 数理計画法による問題解決法, 日科技連出版社.
- 新村秀一(2011c). 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.