Special Thanks to Dr. Yamauchi

イノベーションマネジメント研究科
技術経営専攻

金融リスク・マネジメント

第9講：信用リスク（3）

東京工業大学イノベーションマネジメント研究科
中川　秀敏
E-mail: nakagawa@mot.titech.ac.jp
Office Hour : 17:00-18:30, every Tuesday, at W9-105

※講義資料は、講義前日の午後5時までにはアップするので事前に
http://www.craft.titech.ac.jp/~nakagawa/dir2/lecture.html#TIT2005_1
から各自でダウンロードして用意すること。

Agenda

- 判別分析とは
- 判別分析モデルの作成
- モデル構築上のポイント
  - データの制約
  - 指標選択
  - モデル評価の規準

デフォルト判別分析の基本事項

- 具体的に財務指標データを利用したデフォルト判別モデルを構築する
  - 要するに財務情報をインプットして、その企業にデフォルト懸念があるかないかを判別するモデルを作る
- 今回はシンプルな「線形判別分析」に基づいたモデルを考える

判別分析とは

- 「判別分析」とは文字通り、観測対象が2つのグループに分けられるときに、どちらのグループに属するか不明の個別対象について、付随する変量のデータを用いて、どちらのグループに属する（可能性が高い）かを判別するための統計分析
  - ここでは企業を「安全グループ」か「デフォルト懸念グループ」かに判別したい

判別分析とは

- 一般的には各企業について

$$x_1, x_2, \cdots, x_m$$

という m 個の変量データがあると考える

  - それぞれは例えば「自己資本比率」「流動比率」「売上債権回転期間」「売上高経常利益率」・・・といった財務指標の決算時点の値と見なす。

判別分析とは

- ここで、企業の信用力の評価値 z が

$$z := \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

という形で表されると仮定する (線形判別分析)

- 各企業の評価値 z を 0 と比較して：

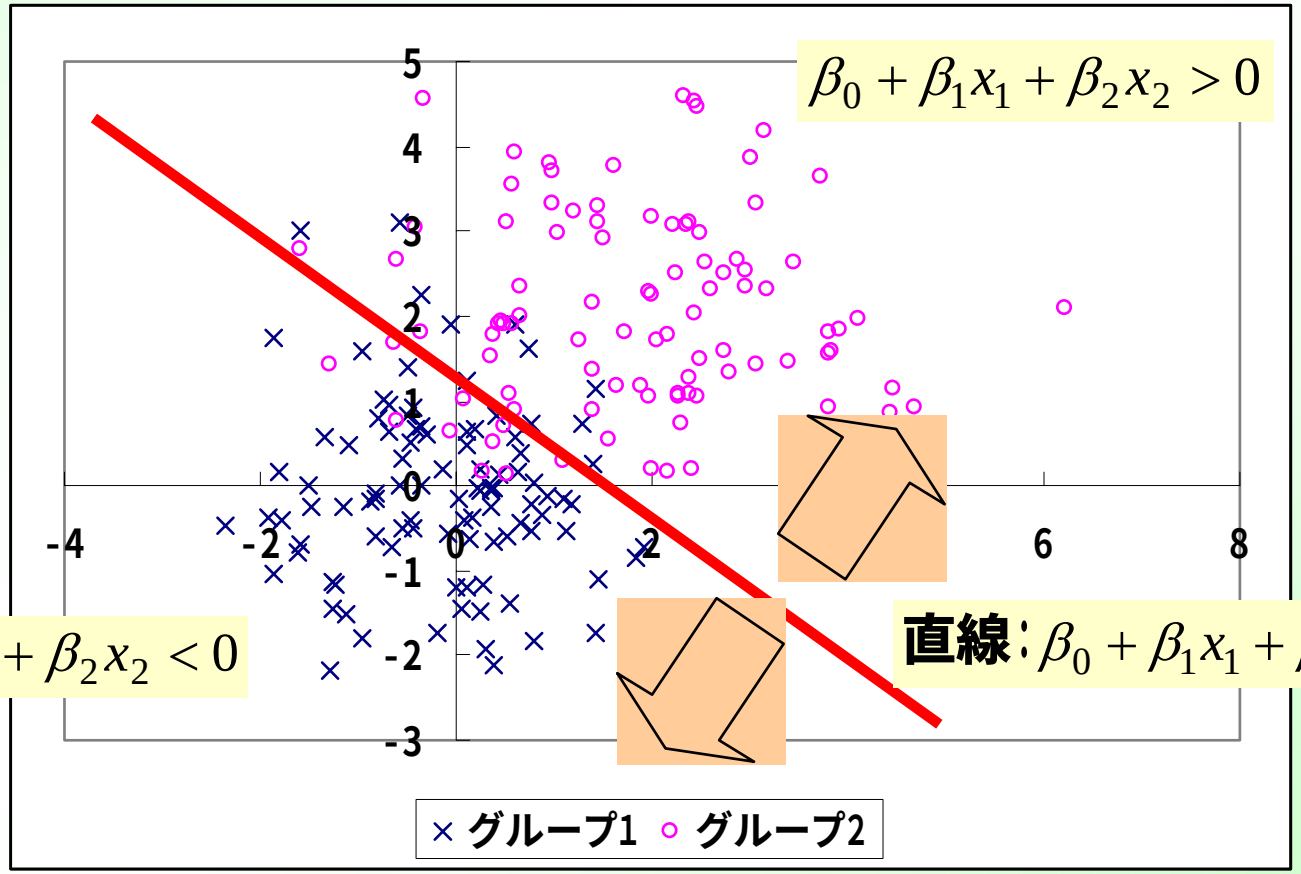
$$z > 0 \Rightarrow \text{安全グループ}$$

$$z \leq 0 \Rightarrow \text{デフォルト懸念グループ}$$

- 問題は　$\beta_0, \beta_1, \beta_2, \cdots, \beta_m$　をどのように決めるか

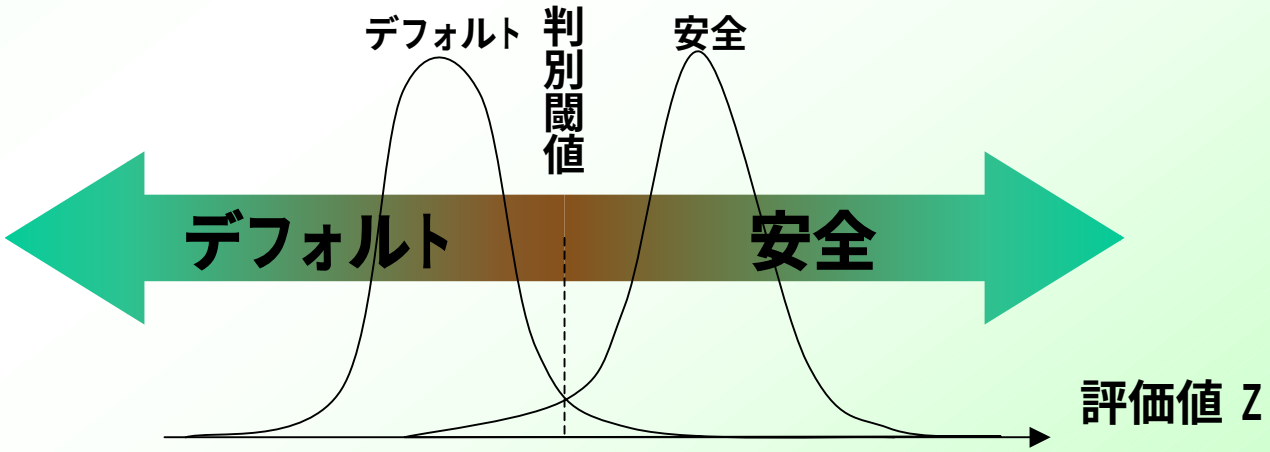
判別分析とは

各要素が2次元の変量データで表現され、以下のような平面に図示される場合、2つのグループを（ある基準に対して）最適に分ける直線を求めることが問題となる。



判別分析とは

- 判別分析を行う上で、本当な必要な前提（デフォルト判別で満たされることはほとんどないが）
  - 各変量（財務指標）の分布は正規分布に従う
  - 両グループの変量の分散共分散行列は等しい
- 結果的に評価値Zは正規分布に従い、デフォルトグループと安全グループは図のように配置されると考えられる



判別分析とは

- Altman の研究の紹介
  - E. I. Altman (1968), “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy”, Journal of Finance, 23, 589-609.
  - 次の5つの指標を用いて線形判別モデルを構築

$$X_1 = \frac{\text{運転資本}}{\text{総資産}},\quad X_2 = \frac{\text{内部留保}}{\text{総資産}},\quad X_3 = \frac{\text{支払利息・税引前当期利益}}{\text{総資産}},$$

$$X_4 = \frac{\text{株式時価総額}}{\text{負債簿価}},\quad X_5 = \frac{\text{売上高}}{\text{総資産}},$$

※) 運転資本 = 流動資産総額 – 流動負債総額,

※) 内部留保 = 税引後当期利益 – 配当・役員賞与

判別分析とは

- Altman の研究の紹介 (つづき)
  - 米国の計66社分のデータ (デフォルトと非デフォルト33社ずつ)
  - 彼の示した結果

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$$

  - 彼はこの評価値を「Zスコア」と呼んだので、このような線形判別関数の値を一般に「Zスコア」と呼ぶことも多い。
  - データセットの抽出方法、財務指標の正規分布性やそれぞれのグループの分散の一致などに言及されていない
    →実際のモデル構築では注意が必要

判別分析とは

- デフォルト判別結果の他への応用
  - 評価値Zを0以上1以下の値に変換して「デフォルト確率」として表現することも行われる
  - また、評価値Zに区切り値を設定し、「△△点以上、○○点未満ならば××格」というように格付を付与することも考えられる

Zスコア	格付け
2.5以上	AAA
1.5～2.5	AA
1～1.5	A
0.5～1	BBB
0～0.5	BB
-0.5～-1	B
-1未満	CCC

判別分析モデルの作成

- 要は企業の信用リスク評価値を

$$z := \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

という形で表すときの係数をパラメータとして、実際の財務データを用いて推定したい。

- 本来は、モデルで利用する財務データとして
  - どのような指標を用いるか？
  - 何個の指標を用いるか？

という「指標選択」が重要だが、ここではそれについては与えられているとして考える。

判別分析モデルの作成

- Zの値が大きいほど信用力が高い（信用リスクが小さい）と考えるとき、

x_i : 値が大きいほど良いとされる指標

$$\Rightarrow \beta_i > 0$$

x_i : 値が小さいほど良いとされる指標

$$\Rightarrow \beta_i < 0$$

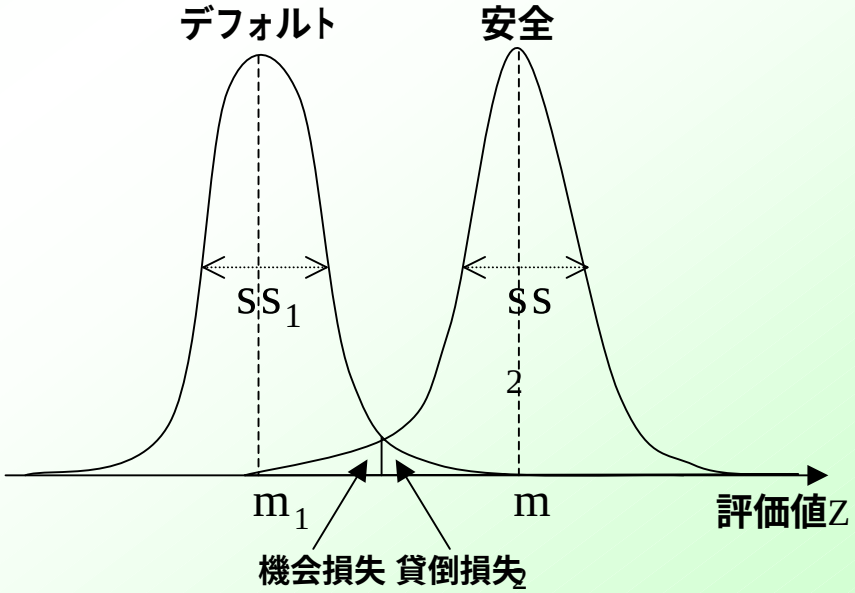
- そもそも、指標の値と企業の評価との間に単調性があることが、線形判別モデルを作る際には必要
  - 「有利子負債÷当期利益」のような指標は単調性があるか？

判別分析モデルの作成

- パラメータ推定は、デフォルトグループと安全グループそれぞれの評価値Zの分布ができるだけ分かれる（重なりが小さくなる）ように行われる

- 一つのアイデアは、以下のGの値が最大になるようにパラメータを決定すること

$$G \equiv \frac{(m_1 - m_2)^2}{ss_1 + ss_2}$$



(参考) 重回帰分析を利用した判別分析モデルの作成

- <u>ある手順</u>のもとでは線形判別分析と重回帰分析の考え方は非常に近く、重回帰モデルの回帰係数を適当に<u>変換</u>すれば線形判別関数が得られる。
- 回帰分析とは「ある変量 y とある変量 $x_1, x_2, \ldots, x_m$ との間に存在する因果関係 (x の変化によって y が変化する、という意味) を明らかにするための分析手法」のこと。
  - 被説明変数 (目的変数) y ： 因果関係の「果」にあたる変量
  - 説明変数 $x_1, x_2, \ldots, x_m$ ： 因果関係の「因」にあたる変量

$$y = \alpha + \beta_1 \times x_1 + \beta_2 \times x_2 + \cdots + \beta_m \times x_m + \text{誤差}\,\varepsilon$$

- 単回帰分析 ： 説明変数がひとつの場合の回帰分析
- 重回帰分析 ： 説明変数が複数の場合の回帰分析

(参考) 重回帰分析を利用した判別分析モデルの作成

①目的変数 y をすべての企業に対して<u>一律</u>、以下のように定める。

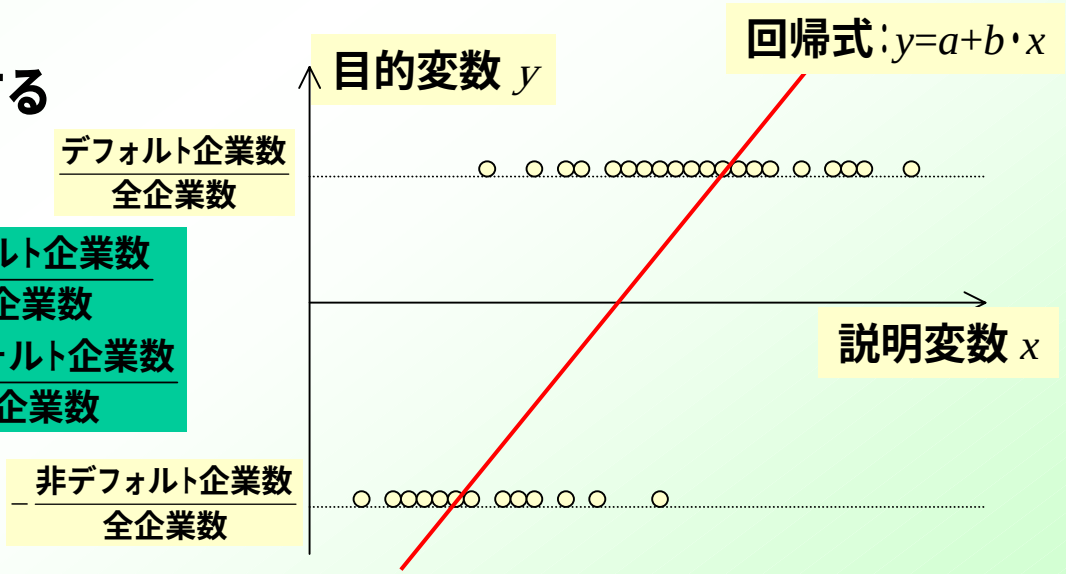
(参考：yの平均値は？)

→ 各企業のデータは、右図の●のように並ぶ

②重回帰分析を実行する

非デフォルト企業 → $\dfrac{\text{デフォルト企業数}}{\text{全企業数}}$

デフォルト企業 → $-\dfrac{\text{非デフォルト企業数}}{\text{全企業数}}$



(参考) 重回帰分析を利用した判別分析モデルの作成

③推定された回帰係数を以下の方法に従って変換する

【準備】

(1) 説明変数 $x_1, x_2, \ldots, x_m$ ごとに、それぞれのグループの平均値を求める。

(2) 説明変数ごとに、「非デフォルトグループの平均－デフォルトグループの平均」を計算し、d_i とおく $(i=1,\ldots,m)$。

(3) 同様に説明変数ごとに、「(非デフォルトグループの平均+デフォルトグループの平均) ÷2」を計算し、q_i とおく $(i=1,\ldots,m)$ 。

(参考) 重回帰分析を利用した判別分析モデルの作成

【変換】

(1) 説明変数 i の判別関数の係数： β_i

$$\beta_i = \frac{(\text{全企業数} - 1)}{\dfrac{\text{非デフォルト数} \times \text{デフォルト数}}{\text{全企業数}} \left\{1 - \left[b_1 \times d_1 + \cdots + b_m \times d_m\right]\right\}} \times b_i$$

(2) 判別関数の定数項：

$$\text{定数項} = -\left[\beta_1 \times q_1 + \cdots + \beta_m \times q_m\right]$$

モデル構築上のポイント

- 財務データを利用したデフォルトの線形判別モデルを実際に構築する際に注意すべき点について言及する
  - モデルの係数推定に用いるデータ
  - 利用する財務指標の種類と数
  - 良いモデルかどうかを見極めるポイント

データの制約

- 実際のデータを用いて

$$z := \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

  の係数 $\beta_0, \beta_1, \beta_2, \cdots, \beta_m$ を推定する。
- しかし、実際の財務データには様々な問題が含まれており、推定の前に準備が不可欠

データの制約

- 財務データの問題
  - 年に1回か数回程度しか分析できない
  - 欠損値、異常値、変則決算（決算期変更）、特定年度の異常（特別利益／損失）
  - 会計制度変更による不連続性（単独→連結、簿価会計→時価会計など）
  - デフォルト事例の少なさ

データの制約

- 財務データの問題の解決策
  - 根本的な解決方法はない
  - 基本的には良質でカバー数の多いデータベースを持つことが重要
  - 現実的には、データ整備に人的リソースをかなり費消する
    - 非常にイレギュラーな状況が多数発生しうるので、機械的に処理するのは困難

指標選択

- どの財務指標を利用して、モデルを構築するかを決める
  - 定性的判断
    - 経験上デフォルト判別に有効と考える指標
    - 指標のタイプ (収益性、安全性、成長性など) の全体のバランス
  - 定量的判断
    - モデルを1つ作成したときに、ある統計量または目的関数値を計算し、それが最適になるような指標の組み合わせを選択する

指標選択

- 定量的判断
  - どのような評価値を考えるかに結果は影響される
  - 線形モデルにおける指標選択
    - 「総当たり法」「変数増加法」「変数減少法」「変数増減法（ステップワイズ）」
    - 自由度調整済み決定係数、AIC（後で詳述）などの基準を最適化

指標選択

- その他の数理モデルを用いた指標選択
  - 決定木(Decision Tree)モデル
    - SAFモデル（白田佳子氏）の指標選択
  - ニューラル・ネットワーク
  - 遺伝的アルゴリズム
    - 複雑な問題の近似解を探す方法として最近注目（山内氏）

モデル評価のポイント

- モデルの適合性
  - モデルがパラメータ推定に用いたデータにどの程度適合しているか
    →残差（＝理論値－観測値）を利用した評価
- モデルの有意性
  - （モデルの適合とは独立して）適用したモデル（説明変数）に意味があるか
    →パラメータについての仮説検定
- 理想としては、「モデルがデータによく適合」し、かつ「モデル自体も有意であること」が望ましい。

モデル評価のポイント

- モデルの適合性（一般論）
  - （モデルから算出される）理論値と観測値の差（残差）に注目して、モデルが観測データにどの程度当てはまりが良いかを調べる。
    - 回帰分析における自由度調整済み決定係数
  - モデルの尤度（比）：推定されたデフォルト確率のもっとらしさ　→　モデル全体での当てはまりの良さを判断する
  - 誤判別率：モデルの判別結果と実際と異なる割合
    → 「機会損失」と「貸倒損失」
  - 個別企業の判別結果

モデル評価のポイント

- 自由度修正済み決定係数
  - 1に近いほどモデルの説明力が高いことを表す
  - 自由度（＝データ数－説明変数の数（定数含む））を考慮することで意味のない変数を追加して、見せかけの説明力を上げることを防いでいる。

$$\overline{R}^2 = 1 - \frac{\frac{1}{N-(p+1)}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\frac{1}{N-1}\sum_{n=1}^{N}(y_n - \bar{y})^2}$$

N：データ数，p：説明変数の数

y_n：観測値，$\hat{y}_n$：モデルによる理論値，$\bar{y}$：観測値の平均

モデル評価のポイント

- 尤度
  - あるモデルを仮定している状況で、その観測データが得られる確率、言い換えると「ある観測データにそのモデルがどれぐらいあてはまっているか」を示す尺度

$\tilde{x}_1, \cdots, \tilde{x}_N$: 観測データ

$f(x;\theta)$: 密度関数（θはパラメータ）

尤度 ＝ $f(\tilde{x}_1;\theta) \times \cdots \times f(\tilde{x}_N;\theta)$

  - 尤度を最大にするようにパラメータを求めることを「最尤推定」という

モデル評価のポイント

- モデルの適合性（一般論）
  - 何個ぐらい変数を取り入れればよいかを判断する尺度
    - 「尤度（もっともらしさ）が同程度のモデルであれば、変数が少ない方がよい」という考え方に基づく
    - 一般に変数の少ないモデルの方が、係数推定リスク（モデルリスクの一種）が低い
    - AIC などの情報量基準を用いて判定

モデル評価のポイント

- AIC(Akaike Information Criteria)

$$AIC = -2(\text{最大対数尤度} - \text{説明変数の数})$$

  - 尤度（もっともらしさ）が大きいほど良いが、変数の数をいたずらに増やすことに対してペナルティを課している。
  - この値が小さいほど良い
    →AICが最小になるような変数の数を求める

モデル評価のポイント

- モデルの有意性（一般論）
  - （モデルの適合性に関係なく）適用したモデル（変数）が意味あるものかどうかを判断する
    - 回帰分析におけるF検定
      →モデルが全体として意味をもつか
    - 回帰分析におけるt検定
      →個々の説明変数が意味をもつか
  - 統計的な検討だけでなく、先に注意をした「係数の符号」が選択した指標と整合的かどうかについても気をつける