

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 R²,
「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

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

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。

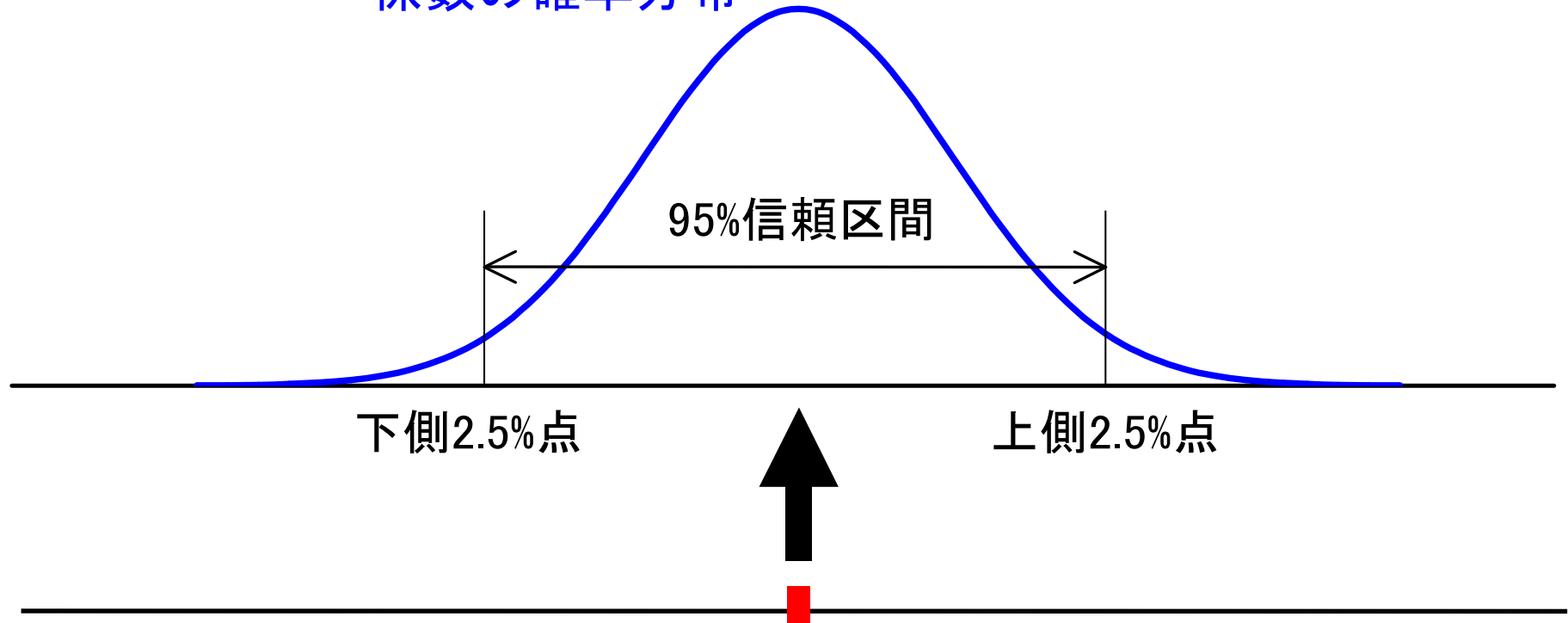
【信頼区間】

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

母集団における
係数の確率分布



標本から計算した係数

★ 95%信頼区間のおおよその値：

$$\underbrace{B}_{\text{係数}} \pm \underbrace{2.0}_{t \text{ 臨} \text{界} \text{値}} \times \underbrace{SE}_{\text{標準誤差}}$$

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

危険率をどこまで下げると
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。