

角度変換

$$\arcsin (y / n)$$

変換してから、分散分析・回帰
(パラメトリックな方法)

割合・比率などの分析に広く使われてきた

もっとも簡単な場合

2つの変数が影響を与える

2独立変数はそれぞれ0と1

2 × 2分散分析

2独立変数の重回帰

		独立変数 1	
		0	1
独立変数 2	0		
	1		

確率が異なる

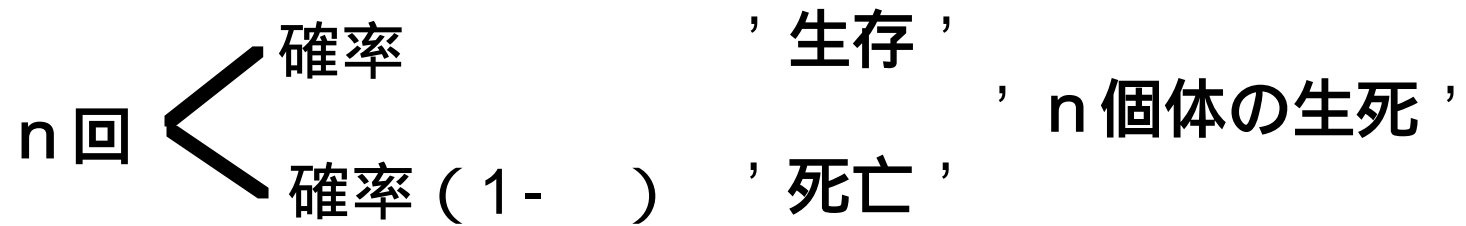
独立変数が2つある 変数の効果間関係

‘関係のモデル’ 2変数が独立に作用する

乗法的（掛け算）	積
ロジスティック的	オッズ比
加法的（足し算）	和

方法

乱数を使って計算



1つの割合 (分母 n)

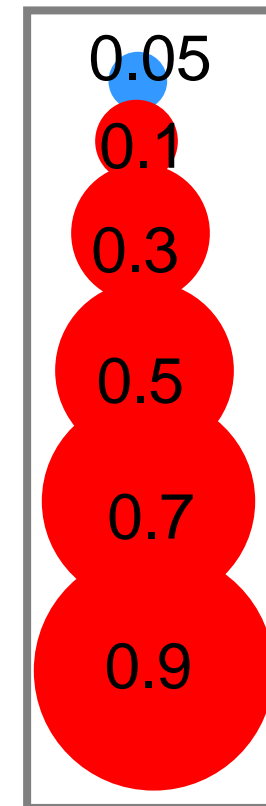
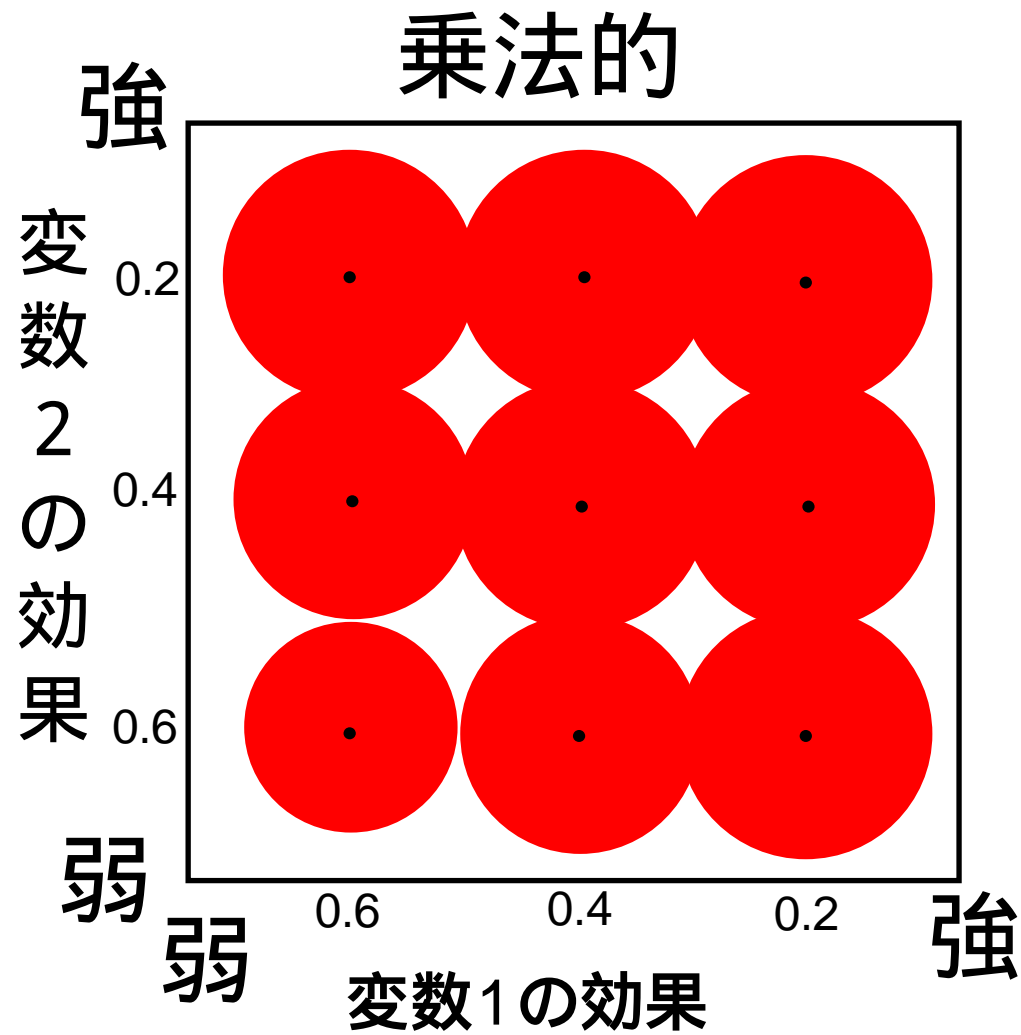
独立変数の組み合わせごとに m 個の割合

分散分析

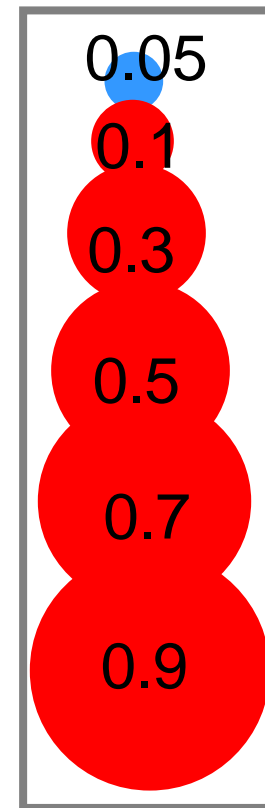
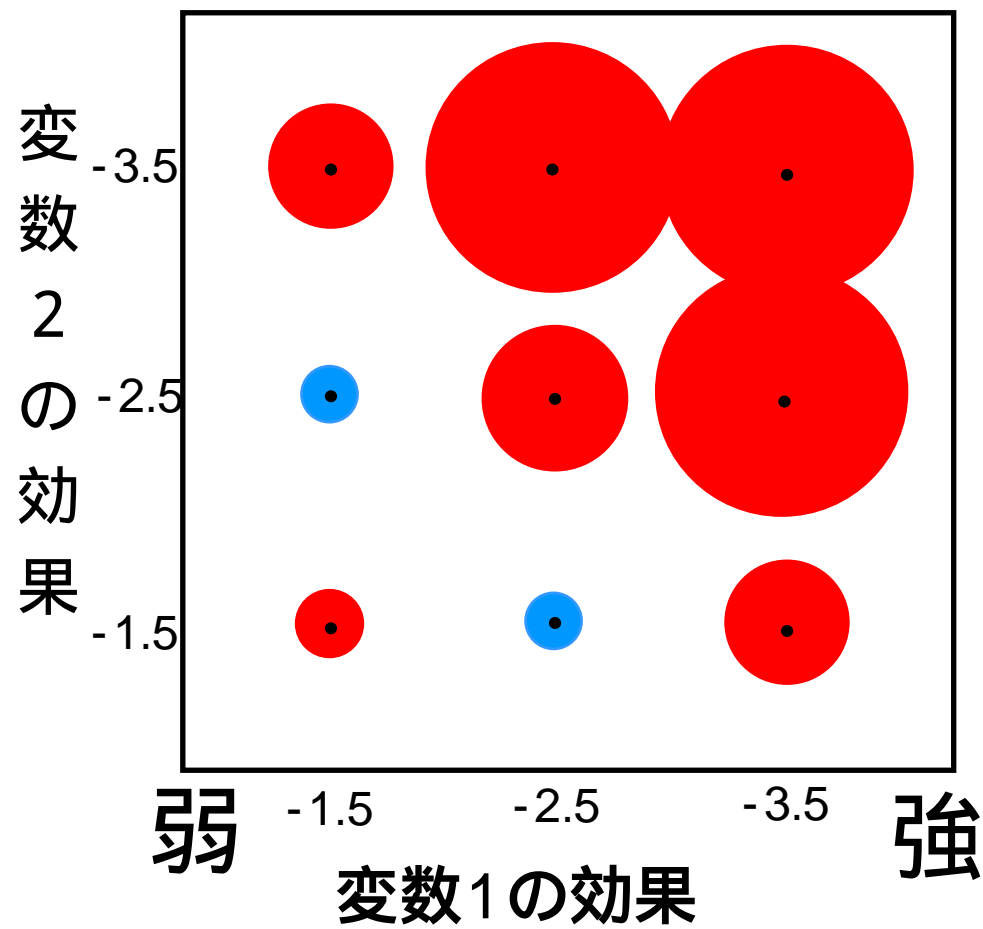
交互作用の検定 (5%水準)

重回帰

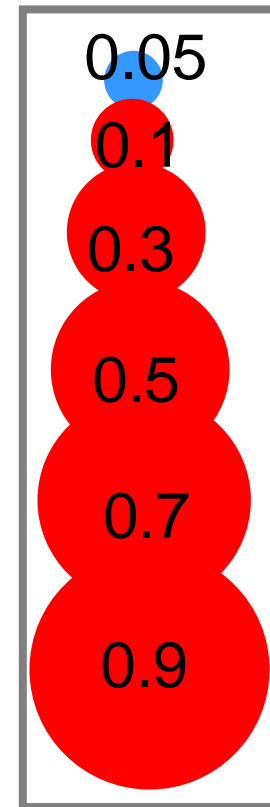
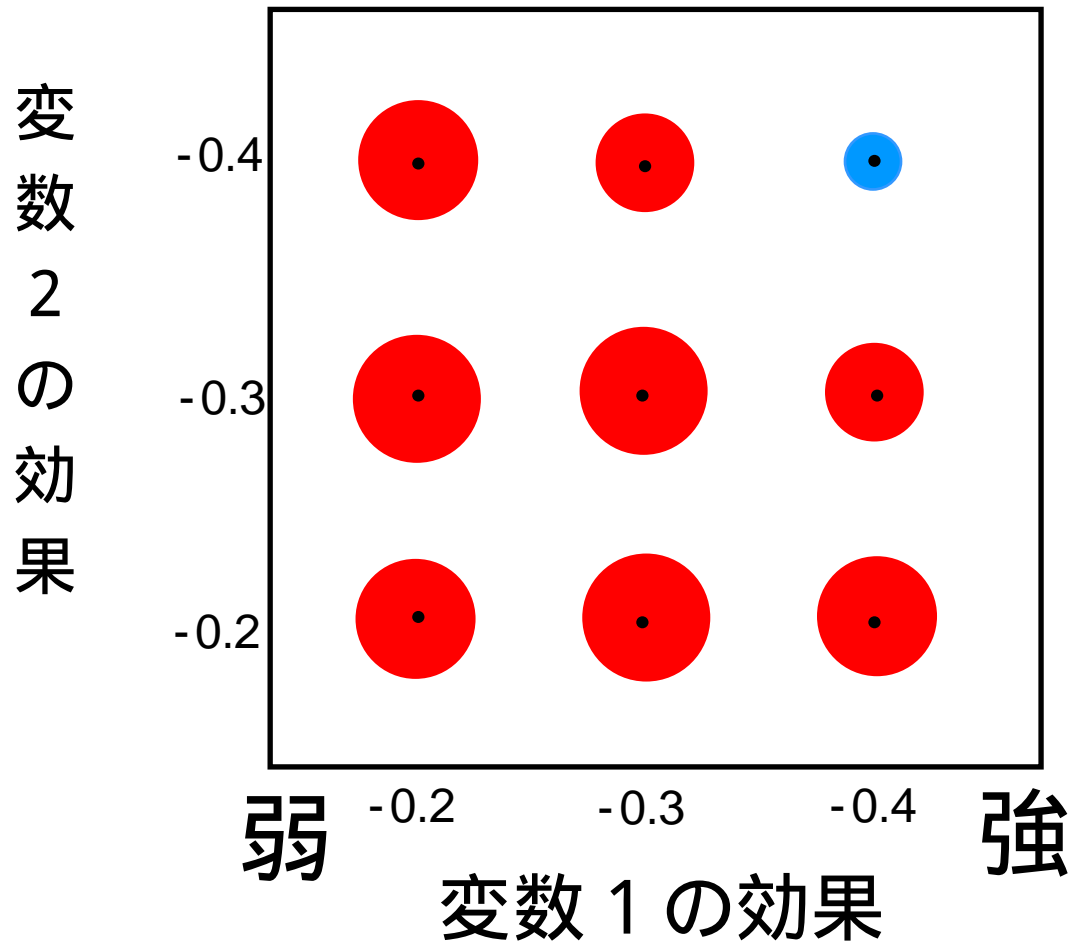
10000回繰り返し



ロジスティック的



加法的

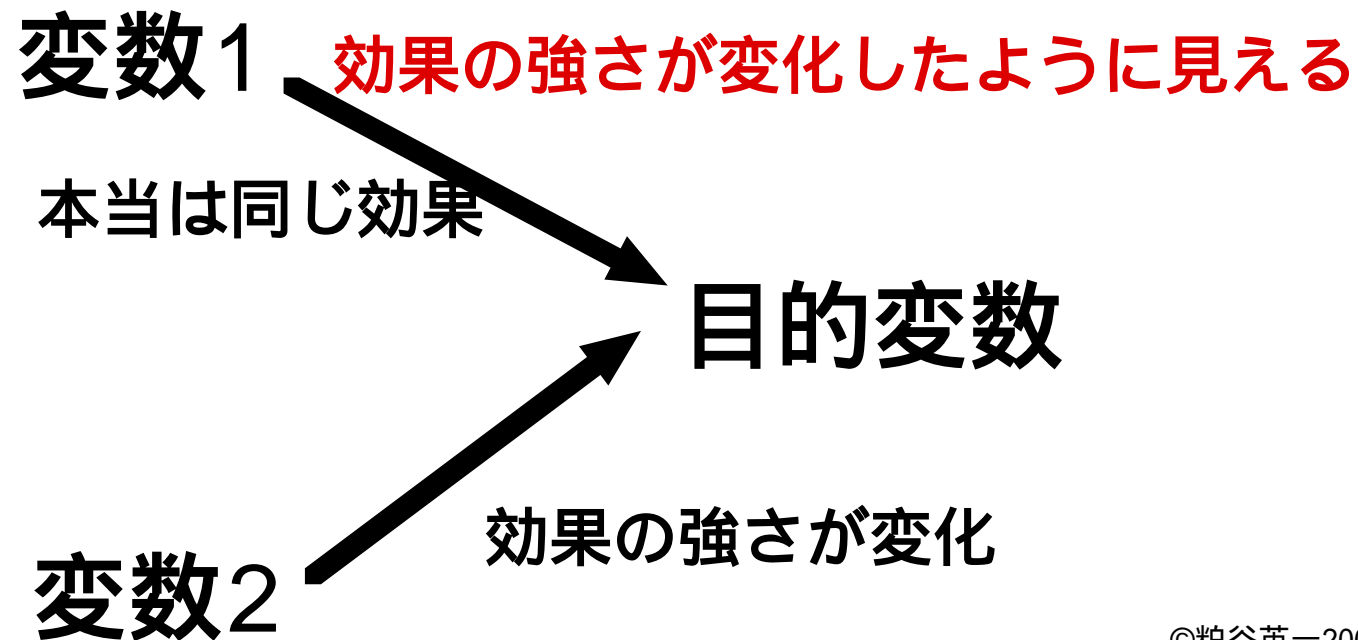


交互作用

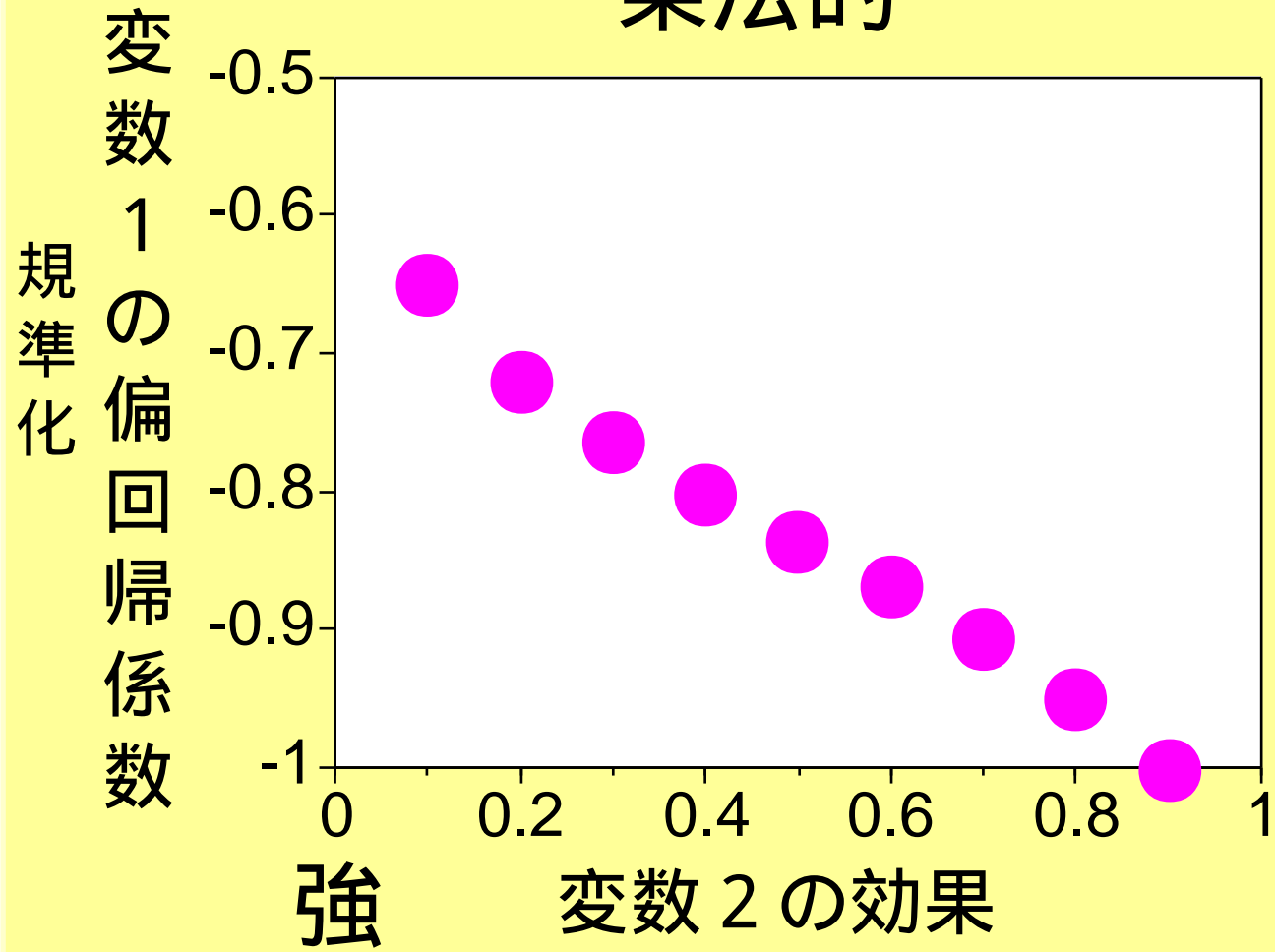
どのモデルでも、独立なはずの変数（要因）間に高率
で有意な交互作用

交互作用があるとは？

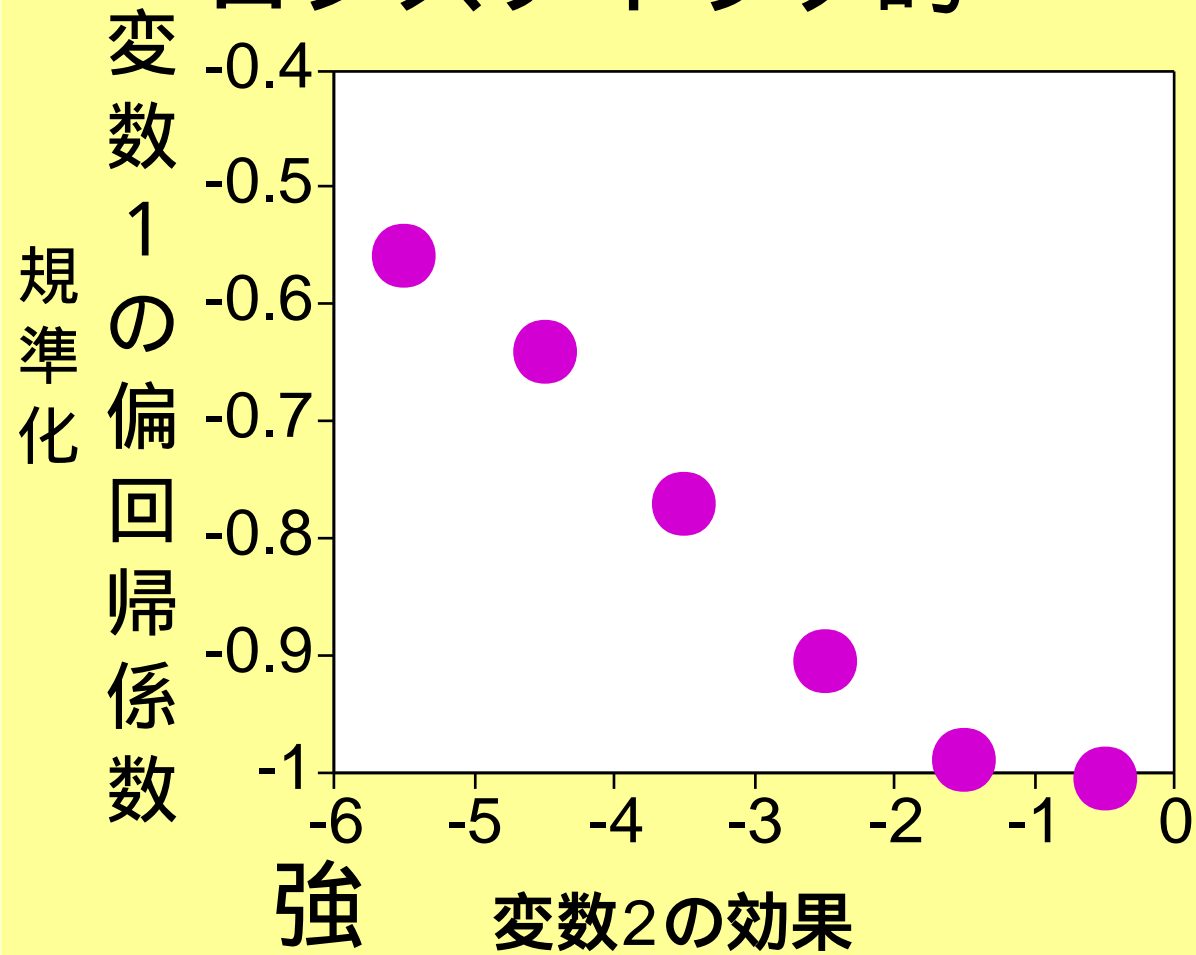
ある変数の効果の強さが
他の変数の効果によってちがう



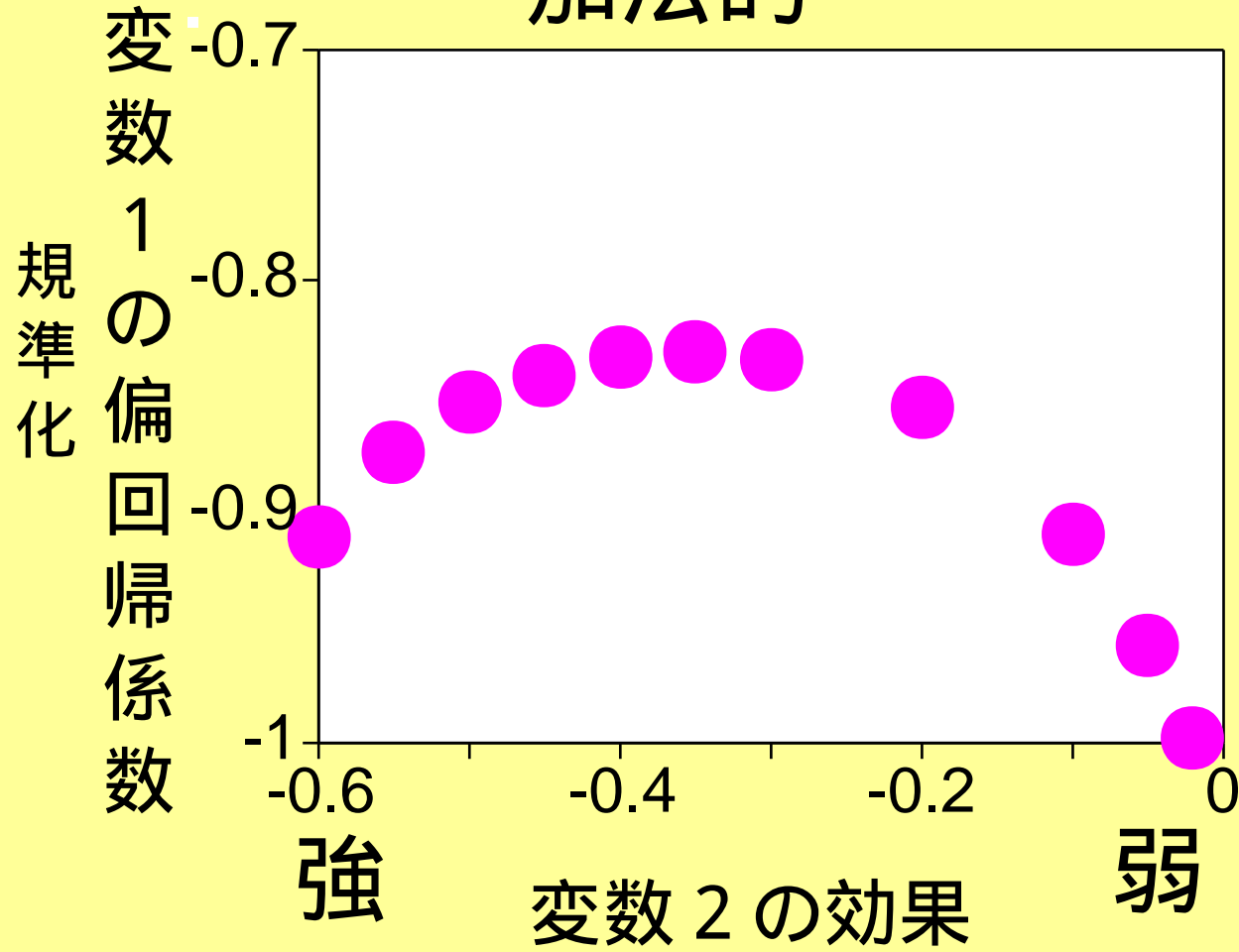
乗法的



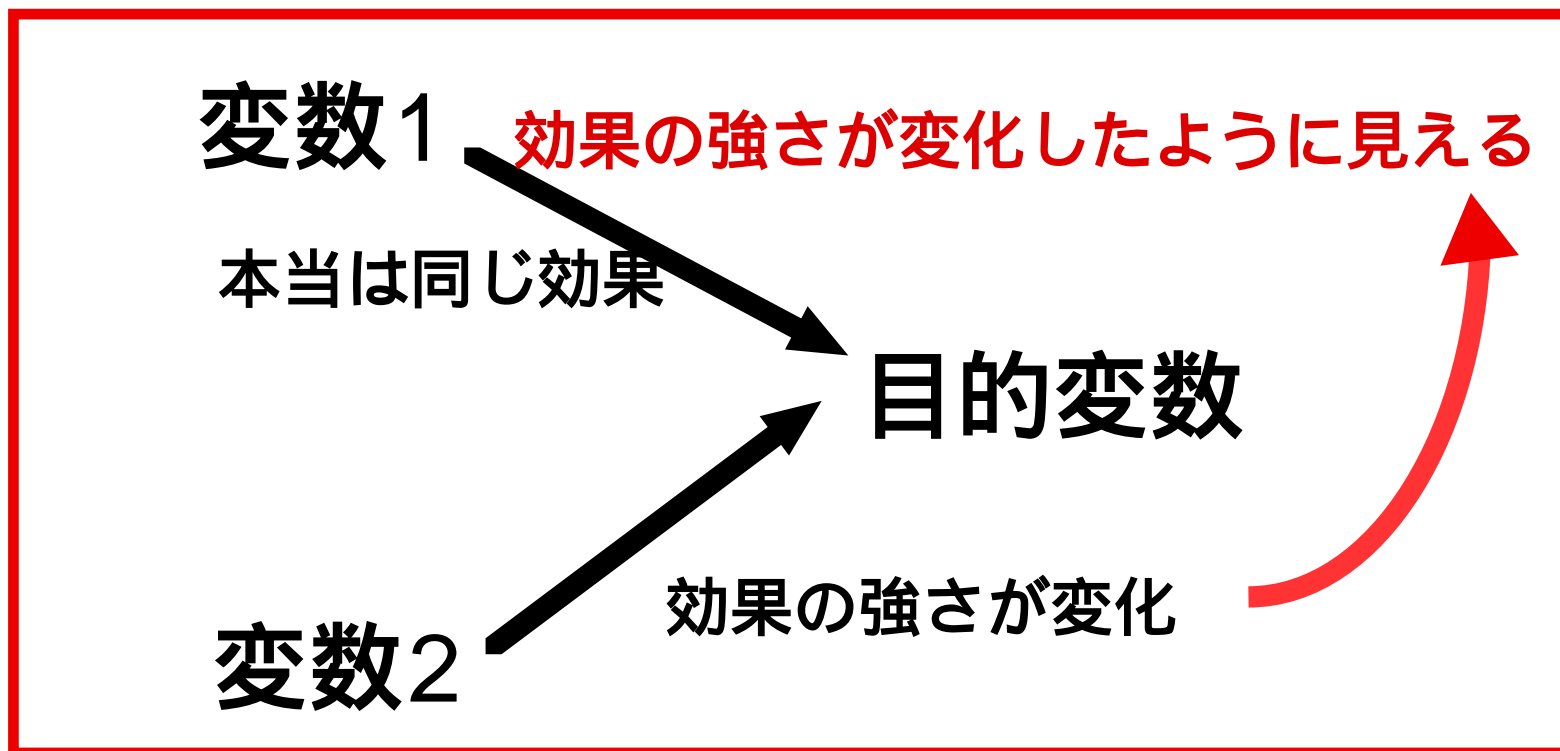
ロジスティック的



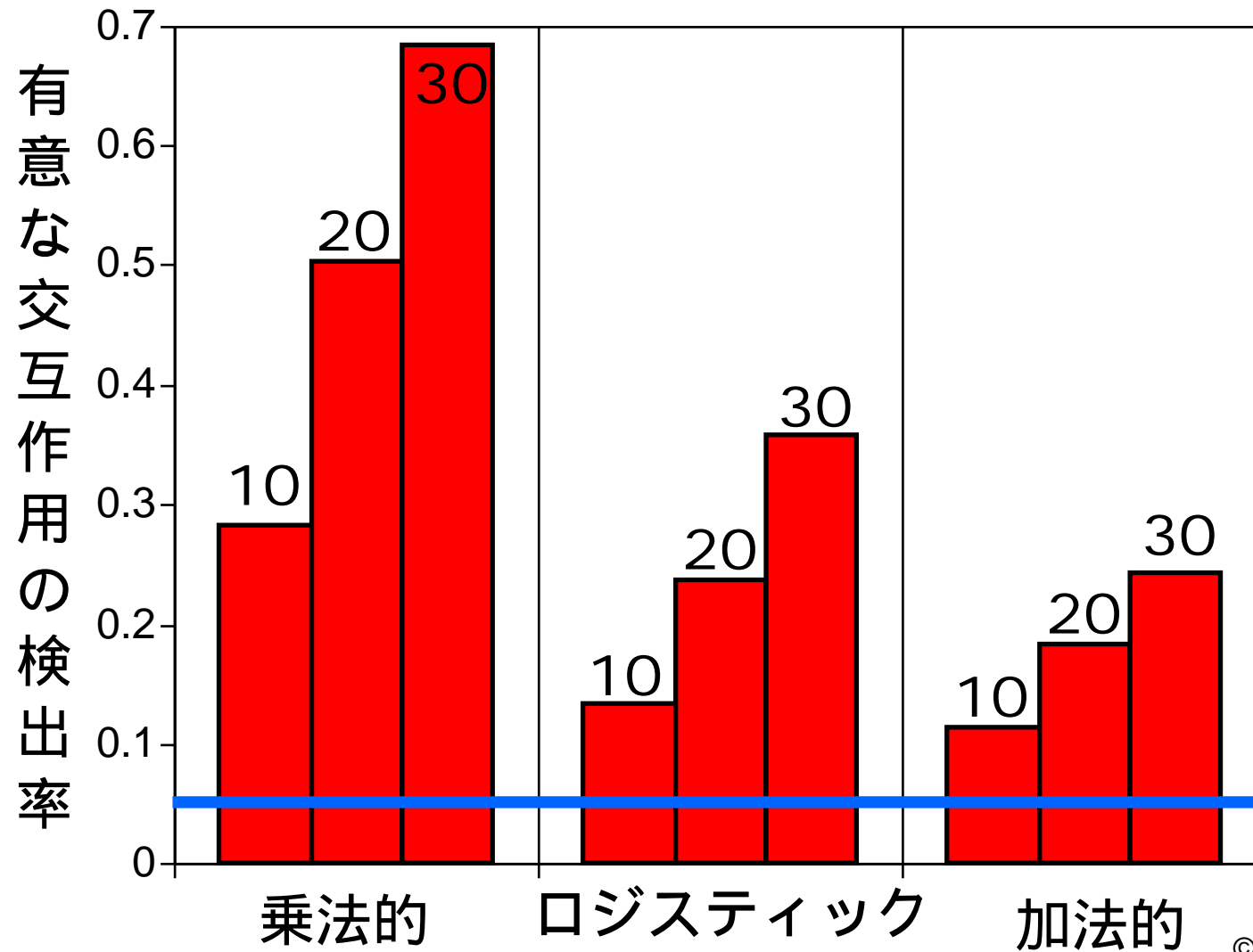
加法的



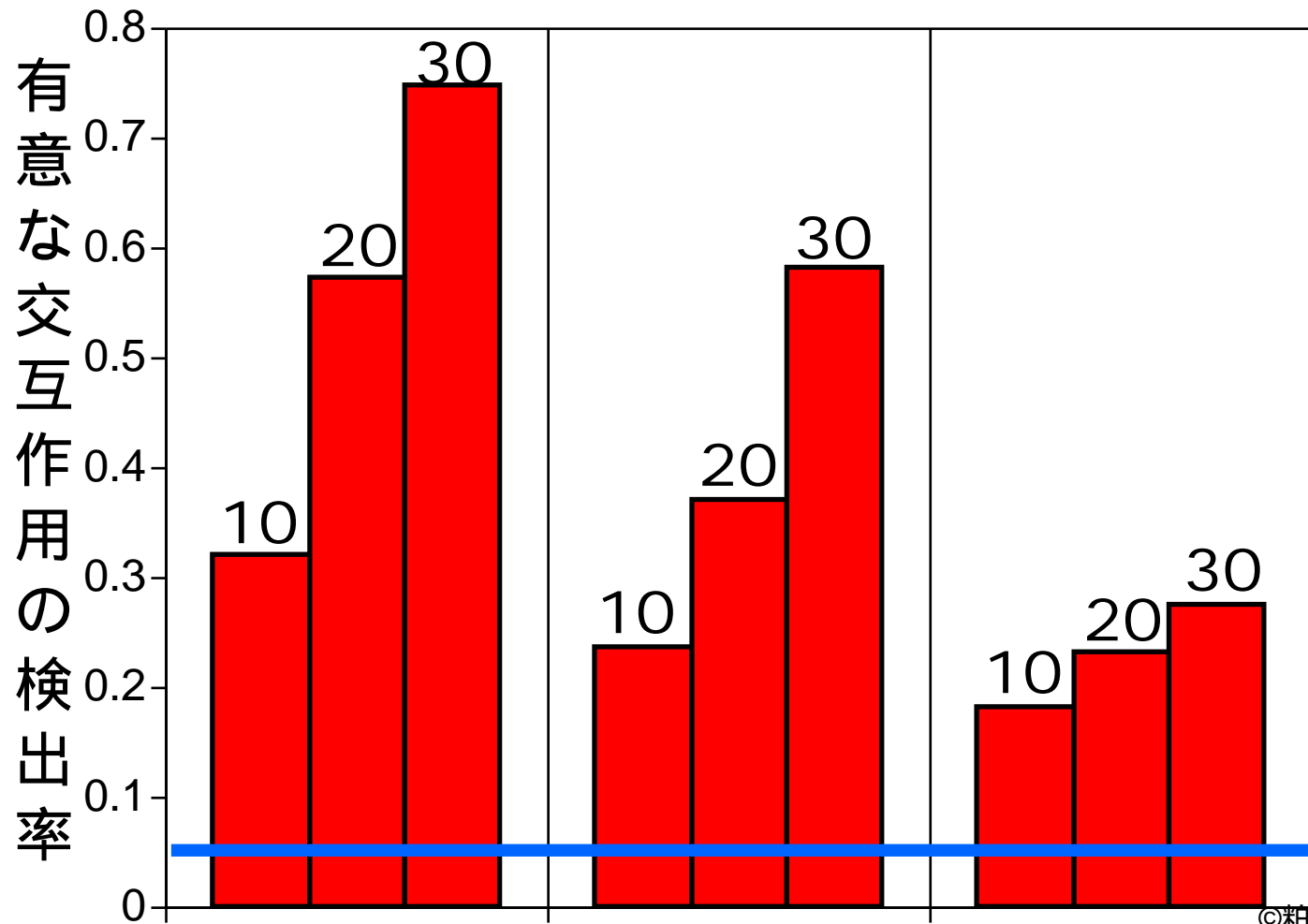
どのモデルでも



割合の個数 (m) が大きいと交互作用の検出率が高い



割合の分母（ n ）が大きいと交互作用の検出率が高い



割合の分母 (n)

割合の個数 (m)

サンプル数が大きいほど影響大

どのモデルでも高率で生じる有意な交互作用

他の変数の影響を受ける偏回帰係数

サンプルサイズが大きいほど悪化

角度変換を重回帰や二元以上の分散分析で使うこと

サンプルサイズを減らせ

変数の効果を小さくしろ（加法的モデルは複雑）

どのモデルでも高率で生じる有意な交互作用

他の変数の影響を受ける偏回帰係数

サンプルサイズが大きいほど悪化

角度変換を重回帰や二元以上の分散分析で使うこと

不適切

角度変換の特殊事情が

べき乗変換

ポアソン変数

単位時間あたりの発生率が一定

起こった回数

ポアソン変数

平方根変換

$$\frac{\sqrt{y}}{\sqrt{y + 0.5}}$$

Freeman - Tukey変換

$$\sqrt{y + 1} + \sqrt{y}$$

対数変換

$$\log(y + 0.5)$$

$$\log(y + 1)$$

方法

乱数を使って計算

ポアソン乱数 母数

独立変数の組み合わせごとにm個のデータ点

分散分析

重回帰

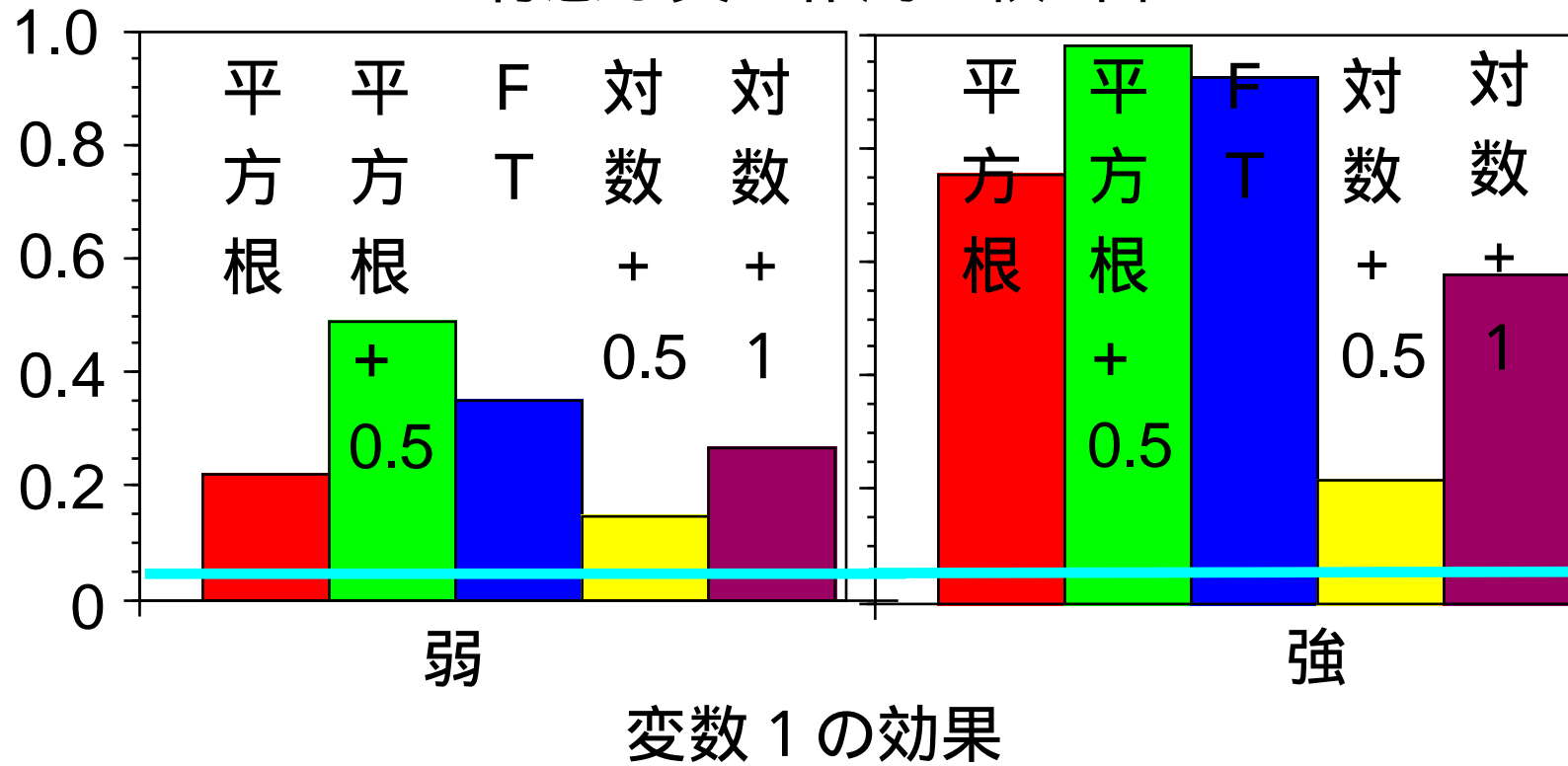
交互作用の検定（5%水準）

10000回繰り返し

関係：乗法、加法

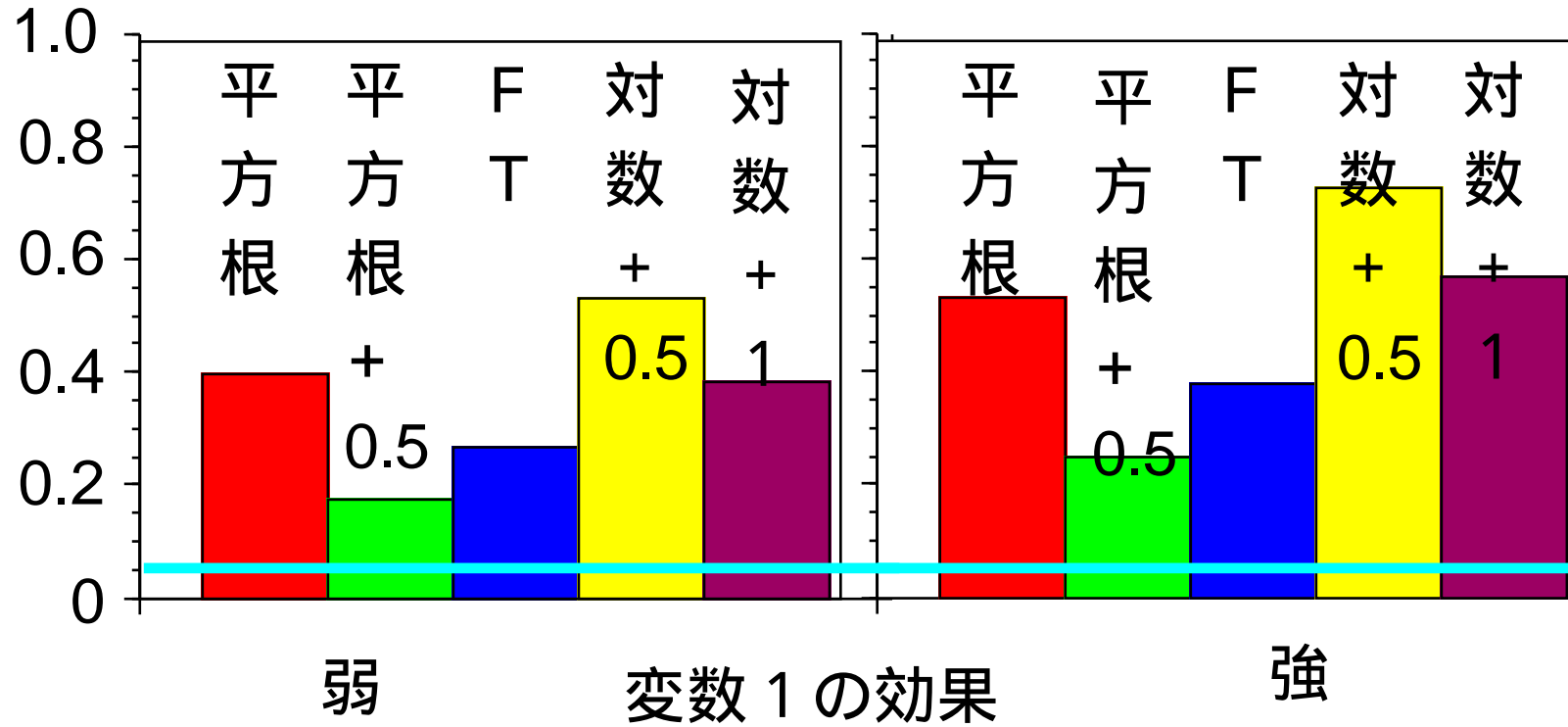
乗法的

有意な交互作用の検出率

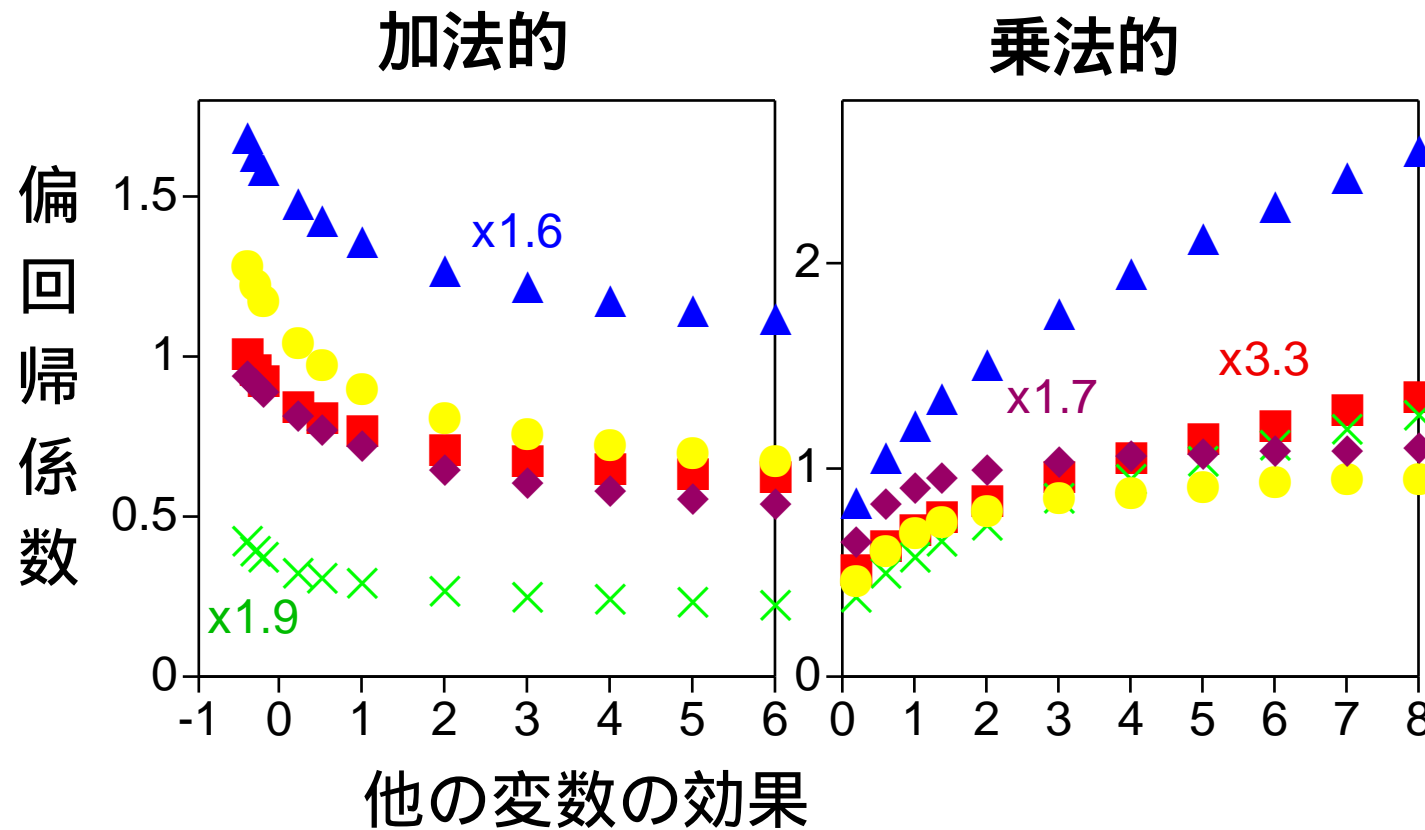


加法的

有意な交互作用の検出率



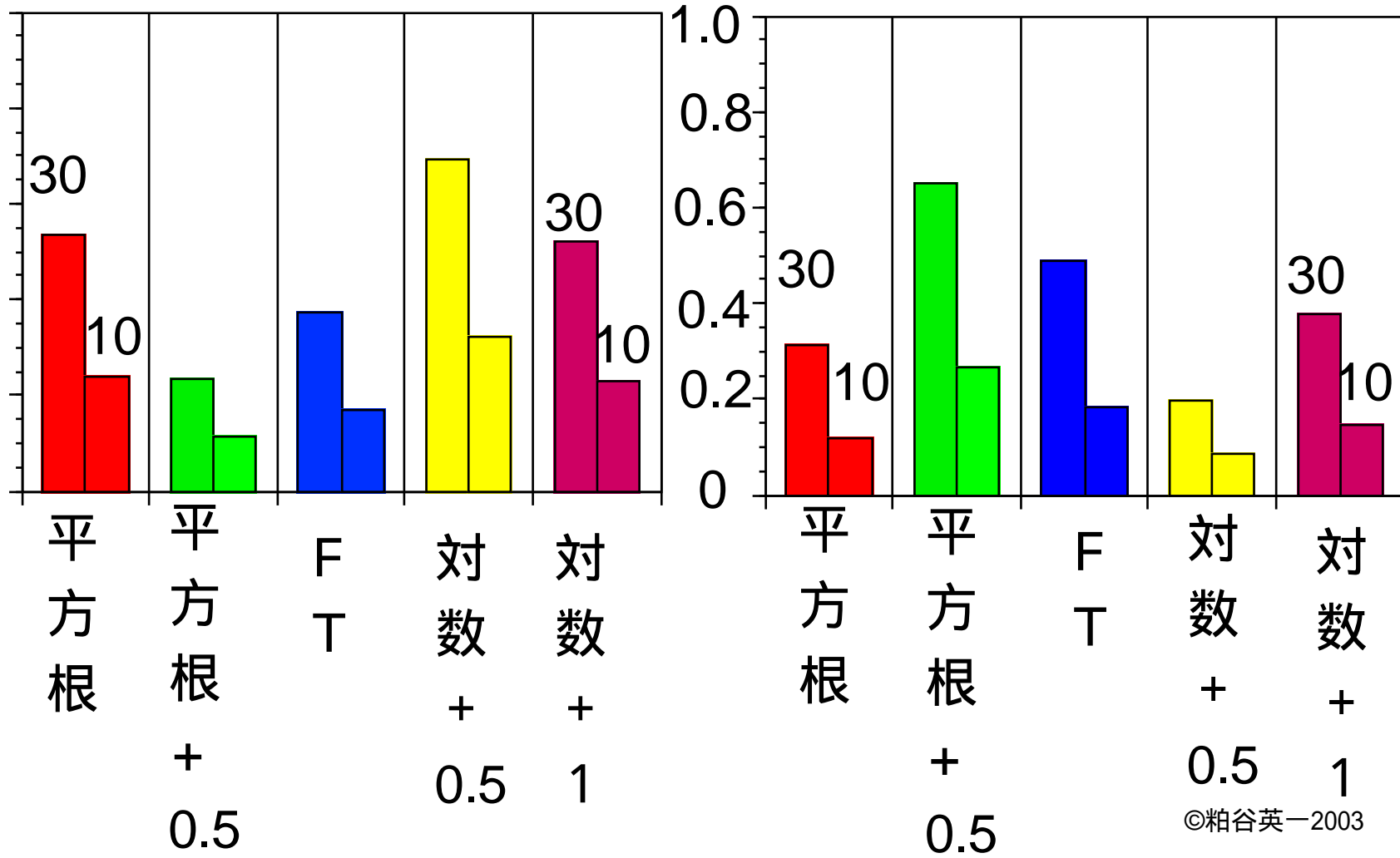
偏回帰係数は他の変数の効果の強さによって変化する



サンプル数が大きいと交互作用は高率で検出される

加法的

乗法的



どのモデルでも高率で生じる有意な交互作用

他の変数の影響を受ける偏回帰係数

サンプルサイズが大きいほど悪化

平方根変換をポアソン変数に対する重回帰や二元以上の分散分析で使うこと

不適切

原因

ばらつきとしての分布を調整しようとする、
同時に期待値にも影響

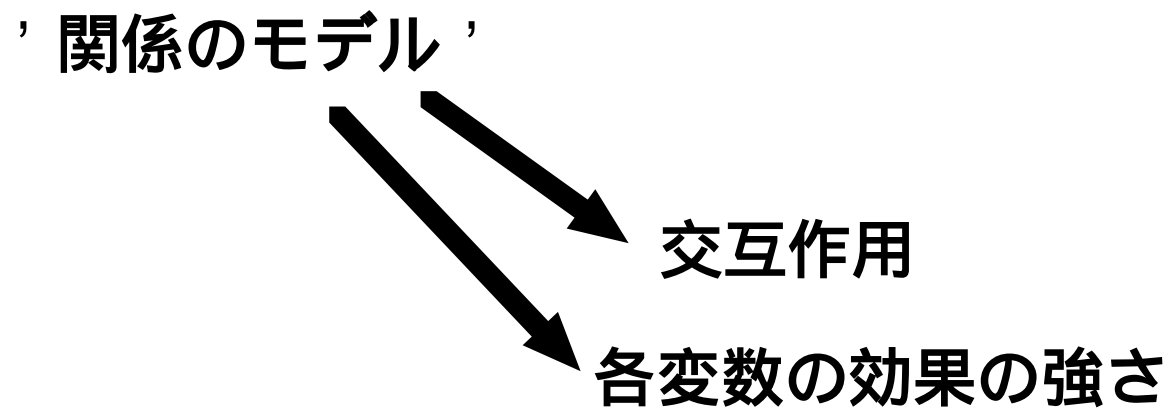
ばらつき（分散）と期待値の関係が両方同時に望んだ
ようになる幸運な場合のみ、複数独立変数の場合に変
数変換は機能する

パラメトリックな方法へのこだわりは無用

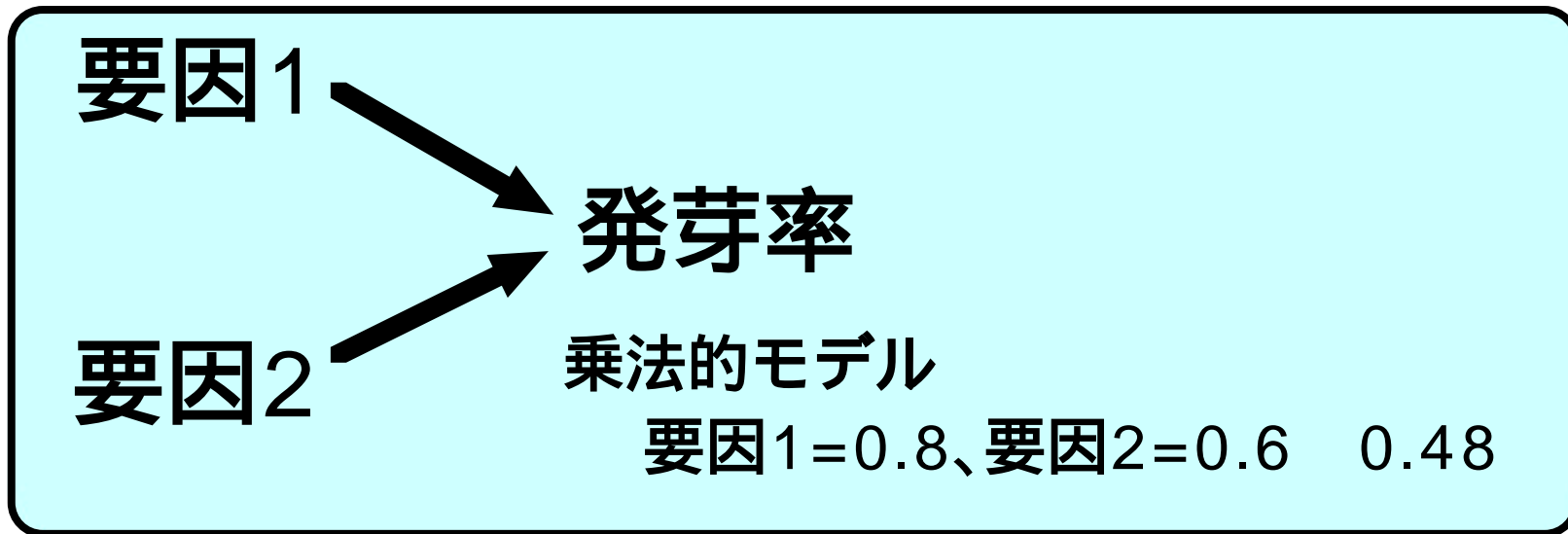
なんでもパラメトリックス

計算量に極端な制約があった時代の名ごり

’ 関係のモデル ’ ↔ 交互作用



(仮想的な例)



誤った結果

加法的モデル
ロジスティックモデル

を想定した統計的方法

誤差すなわちばらつき

あるタイプのデータ

誤差の分布をはっきりさせる



こういう研究への過小評価

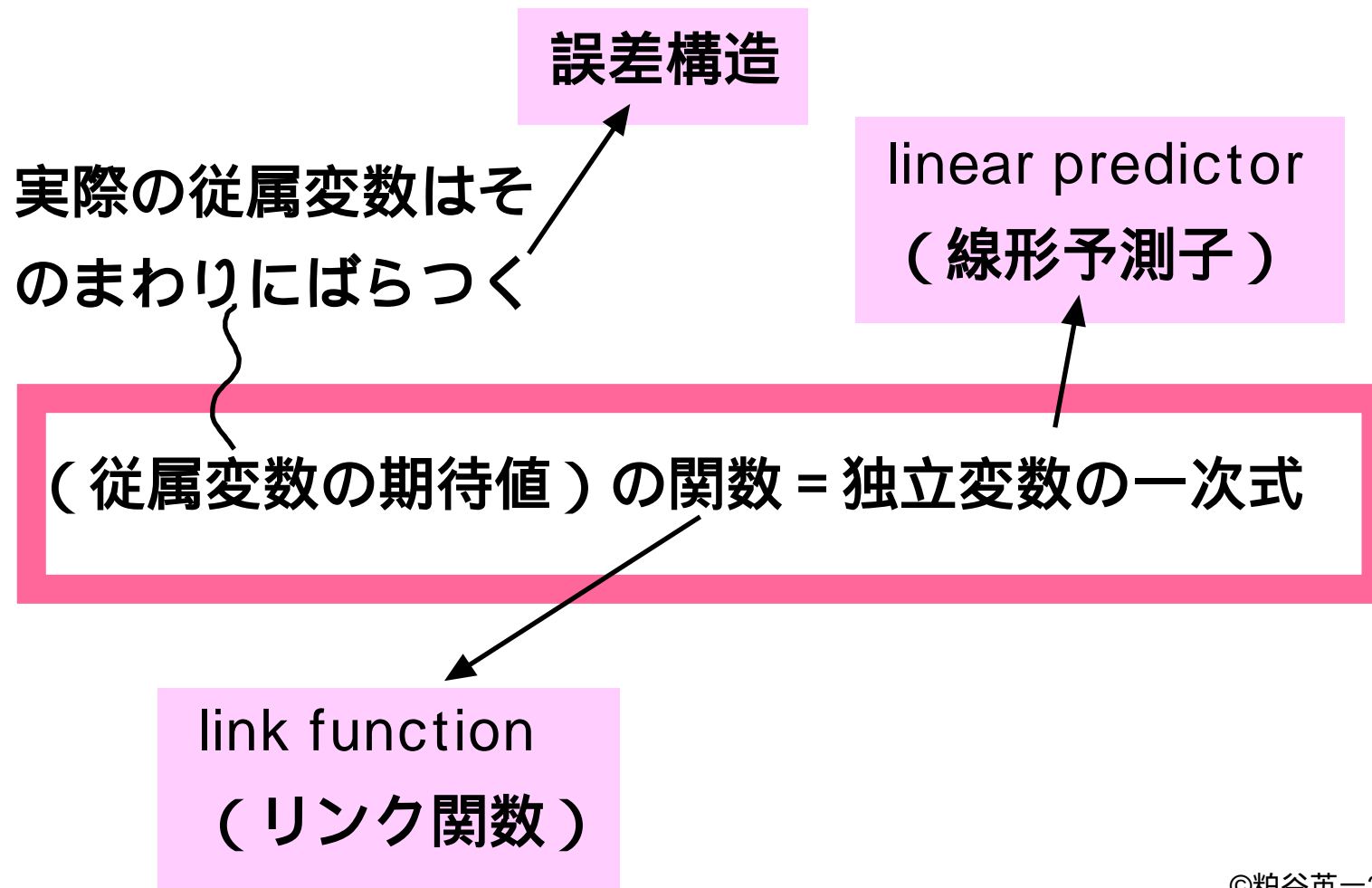
対策

原因

ばらつきとしての分布を調整しようとする、
同時に期待値にも影響

ばらつきとしての分布と期待値を分離する

一般化線形モデル (Generalized Linear Models, GLM)



期待値の関係



link function
(リンク関数)

ばらつき



誤差構造

重回帰
分散分析
ロジスティック回帰
分割表

そのまま

そのまま

ロジット

対数

等分散の正規分布

等分散の正規分布

二項分布

ポアソン分布

ばらつきとしての分布

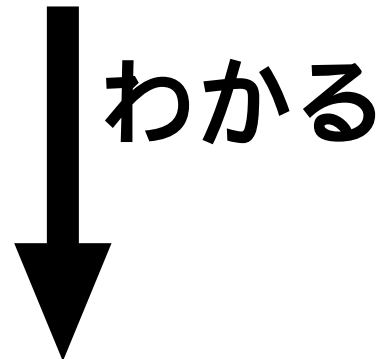
期待値との関係

分離するためには

期待値との関係 独立性とは何かはっきりさせる
実証的な関係の検討

ばらつきとしての分布 データの量

従属変数の期待値と独立変数の関係
誤差分布



最尤法（含、GLM）

ソフトウェア