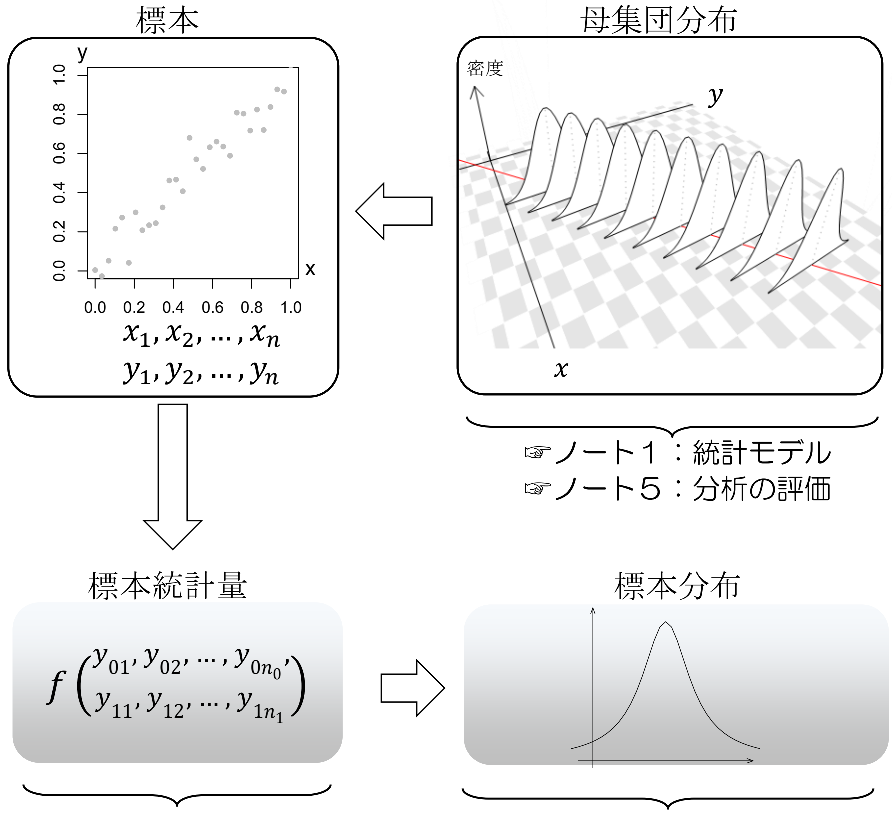
第5講 重回帰分析

学習の目標

- ☐ 回帰モデルにおいて独立変数の数が複数含まれるものを重回帰モデルということが分かる。
- ☐ 回帰モデルを数式や図によって表現することができる。
- ☐ 研究者の興味や仮説に応じて母集団において柔軟なモデルを立てる意義ややり方についての基礎が理解できる。
- ☐ ある独立変数が別の変数を経由して従属変数に間接効果を持つとき仲介をなす変数を媒介変数ということが分かる。
- ☐ 複数の要因が連動して変化するためどちらに影響したか判断できない状況を交絡と呼ぶことが分かる。
- ☐ 調整変数の存在によって独立変数単体の主効果では説明できない交互作用効果が生じることがあることが分かる。
- ☐ 独立性の仮定を乱すクラスターを持つデータ構造に対して階層モデルを想定する必要性が理解できる。
- ☐ 重回帰モデルの各独立変数にかかる係数を偏回帰係数と呼び、その解釈の仕方を理解することができる。
- ☐ 独立変数にかかる係数の大きさを比較する手段として標準偏回帰係数を用いることの意義が分かる。
- ☐ 推定した回帰モデルのデータへの適合度を測る基準として重相関係数と決定係数を用いることができる。
- ☐ 重回帰式の検定および各偏回帰係数の検定、そして、信頼区間や予測区間についても理解をし、使用できる。
- ☐ 過学習、解釈可能性、多重共線性などの問題が懸念される際に、独立変数を選択しモデルの改良を行うことができる。
- ☐ 罰則を設けた最小二乗推定について理解することができ、ラッソ回帰を用いたモデル縮約を行うことができる。
- ☐ モデル比較の必要性を理解し、情報量基準やクロスバリデーションを用いてモデルを選択することができる。
- ☐ 多重共線性に解決の一つの対策として、主成分分析を援用し独立変数間の相関をゼロにしたモデルを提案できる。

見取り図



☞ノート１：統計モデル
☞ノート５：分析の評価

☞ノート２：統計量の基準
☞ノート３：回帰係数の点推定
☞ノート４：回帰係数の検定・区間推定

☞ノート４：t 分布

データの形式

ID	予測変数	応答変数
1	0.3	2.1
2	0.1	3.2
⋮	⋮	⋮
n	1.2	1.5

（1）目的（リサーチクエスチョン）

標本が採られてきていると想定される二種類の母集団の母平均（期待値）に差があるかのか否か、判断を下す。

（2）考え方

標本の平均値の差の大きさが、標準誤差何個分なのかを計算して、平均値差の大きさを判断したいのだが、それは標本をものすごくたくさんとってヒストグラムを作らないと無理。

そこで、今回一回きりの標本から標準誤差の推定値を作り、「標本の平均値の差の大きさが、推定された標準誤差何個分なのか」を計算し、それを t 値と呼ぶ。「二つの母集団平均は完全に同じだ」という仮説が正しい時 t 値が従う標本分布が t 分布である。

t 値がこの t 分布のどこに位置しているのかを調べ、もし端の方に位置しているのなら「ありえないこと」が起きたと考え先の仮説を棄却する。

📖 ノート１　母集団に対する仮定：基本的な統計モデル

（１）統計モデル

母集団に対する統計モデルを少しずつ複雑なものにしていくことで、細やかなモデルを提案することができる。

例１：二群の差の検定（対応なしのｔ検定）

数式の上では全く同じ構造

ダミー変数

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

線

例２：単回帰分析

これが線を生む

$$y_i = \beta_0 + \boxed{\beta_1 x_i} + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

面

例３：重回帰分析

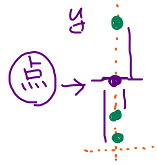
これが線を面にする!

$$y_i = \beta_0 + \beta_1 x_{1i} + \boxed{\beta_2 x_{2i}} + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \boxed{\beta_3 x_{3i}} + \cdots + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

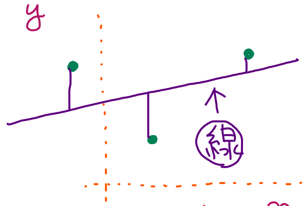
Ⓟ 点 / 線 / 面と平均

(1) 平均　(第二講)



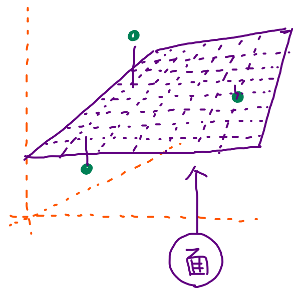
$y_i = \beta_0 + \varepsilon_i$
$\varepsilon_i \sim N(0, \sigma^2)$

(2) 単回帰(第四講)



(対応なしt検定:第三講 x)

(3) 重回帰(第五講)



Ⓟ 最小二乗基準

上の(1)〜(3)は全て、
データとの「二乗距離」を
最小化に見つける
点 / 線 / 平面なのだ!

Ⓟ 添え字

独立変数を区別
するために、添え字を
用いる。

x_{1i}：第1独立変数の
i番目の値

x_{26}：第2独立変数の
6番目の値.

Ⓟ パスモデルの慣習

(1) 変数

- 独立変数
- 従属変数
- 誤差変数.

□：観測変数

○：潜在変数

(2) パス (Path)

左辺の変数
(予測される側)

↑

右辺の変数
(予測する側)

(3) 分散

その変数の分散

共分散

2つの独立変数の共分散

Ⓟ 分散共分散行列

	x_1	x_2	x_3
x_1	$S^2_{x_1}$	$S_{x_1x_2}$	$S_{x_1x_3}$
x_2	$S_{x_1x_2}$	$S^2_{x_2}$	$S_{x_2x_3}$
x_3	$S_{x_1x_3}$	$S_{x_2x_3}$	$S^2_{x_3}$

→ 代わりに相関係数

$r_{x_1x_2}$ などといわれたら、

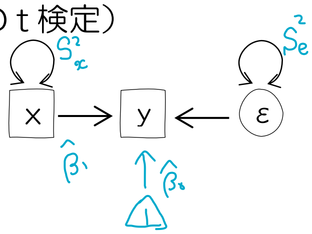
相関係数行列という.

（２） 図による表現１ ：パスモデルとパス解析

これは、変数の関係を中心に発想する視覚的表現。

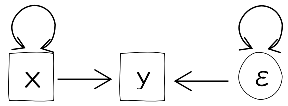
例１ ：二群の差の検定（対応なしのｔ検定）

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$



例２：単回帰分析

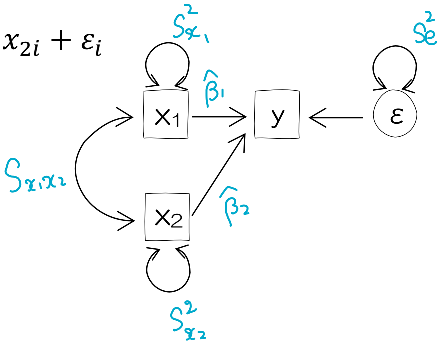
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$



例３：重回帰分析

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

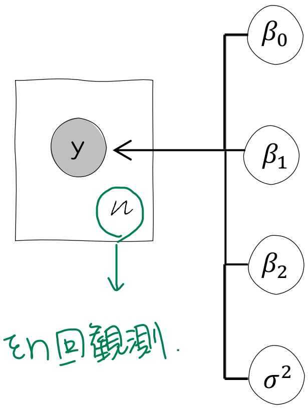
	x_1	x_2
x_1	r_{11}	r_{12}
x_2	r_{21}	r_{22}



（３） 図による表現２ ：グラフィカルモデリング

これは、パラメータの関係を中心に発想する視覚的表現。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

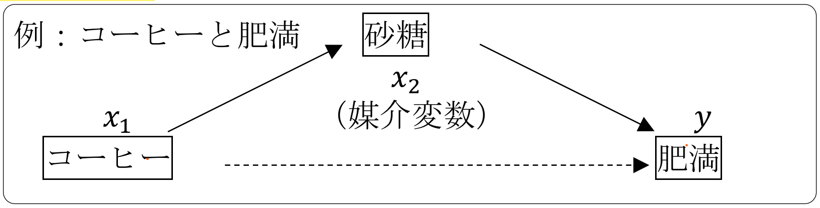


yをn回観測.

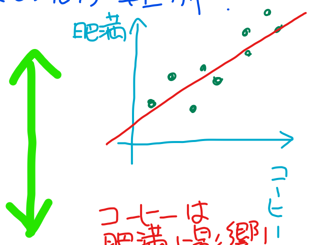
📖 ノート２ †母集団に対する仮定：発展的なモデル

Ⓟ媒介モデル

（１）媒介モデル Mediation Model



(モデル1) 単回帰モデル



コーヒーは肥満に影響!

(モデル2) 媒介モデル

① コーヒー自体は肥満に無関係

② 砂糖の効果(大)

① 直接効果 Direct effects

独立変数から従属変数への直接的な影響の指標。

② 間接効果 Indirect effects

ある独立変数から別の変数（媒介変数／中間変数）を経由して従属変数に伝わる影響のこと。

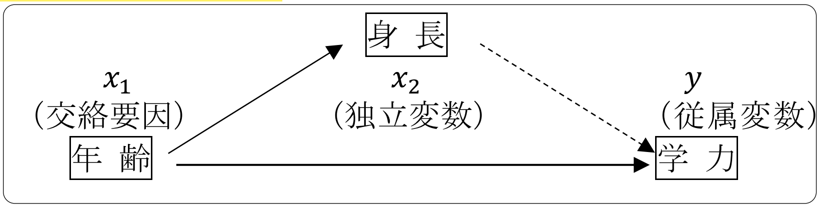
(パート1) $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ ← 単回帰モデル

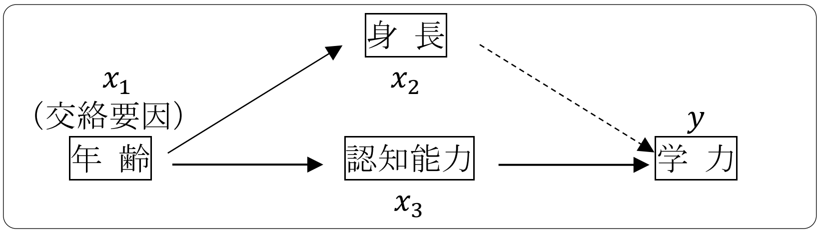
(パート2) $x_{2i} = \gamma_1 x_{1i} + u_i$ ← 媒介

x_2がx_1に影響を受けている

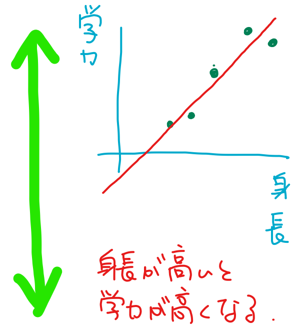
（２）交絡要因／共通因子を持つモデル　（☞ 多重共線性）

Ⓟ 交絡要因





(モデル1) 単回帰



身長が高いと学力が高くなる。

(モデル2) 交絡要因あり

① 年齢がupすると、学力は高くなる。

② 年齢を考慮に入れると、身長の影響はない

① 交絡 Confounding

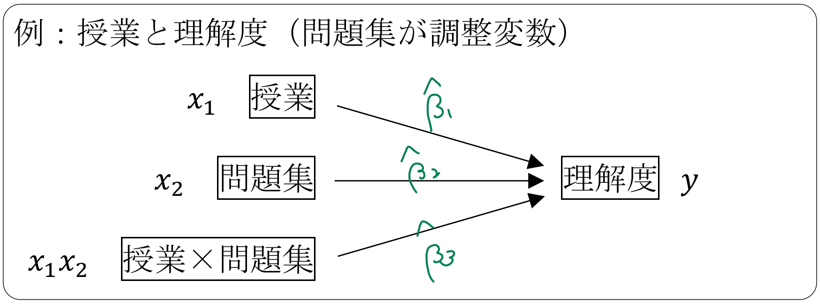
二つ以上の要因が連動して変化するため、そのうちどれが結果に影響したかが判断できない状態になること。

② 交絡変数 Confounding variables

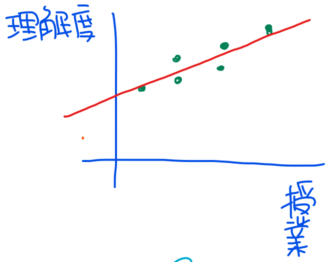
独立変数x_2の上流側にある、独立変数x_2と目的変数yの両者に影響をもたらす要因のこと。

Ⓟ交互作用モデル

（3）交互作用モデル　Interaction Model

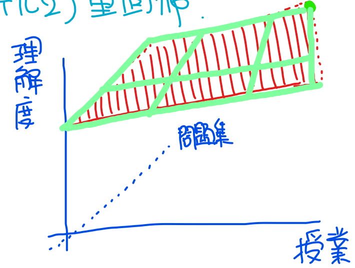


(モデル1) 単回帰



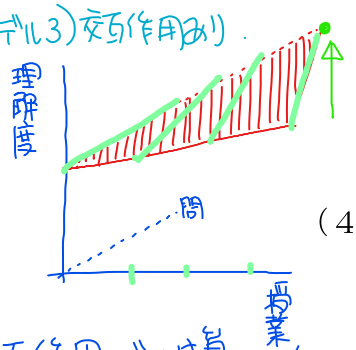
① 調整（抑制）変数　moderator (suppressor) variable
x_1とyの関係が第三の変数x_2の値によって変化するとき、このx_1と交互作用を持つx_2を調整変数という。

(モデル2) 重回帰



② 交互作用　Interaction
二つ以上の変数が組み合わさって生まれるもともとの変数単体の効果（主効果）では説明できない効果のこと。

(モデル3) 交互作用あり



(パート1)　$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$　← 重回帰

+) (パート2)　β_2がx_1の値で変化 →　$\beta_2 = \gamma_0 + \gamma_1 x_{1i}$ ← 交互作用

$y_i = \beta_0 + \beta_1 x_{1i} + \gamma_0 x_{2i} + \gamma_1 x_{1i} x_{2i} + \varepsilon_i$

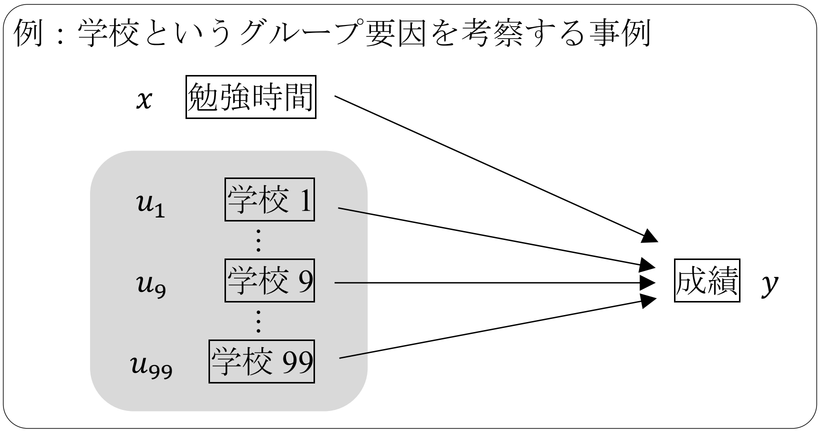
Ⓟ交互作用がかけ算のワケ

$y_i = \beta_0 + \beta_1 x_{1i} + \gamma_0 x_{2i} + \gamma_1 x_{1i} x_{2i} + \varepsilon_i$

$= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i$

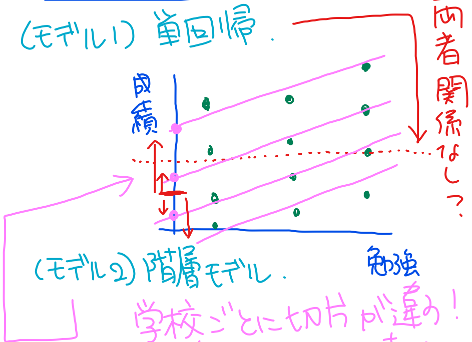
$= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 \underbrace{x_{1i} x_{2i}}_{\text{かけ算}} + \varepsilon_i$

（4）階層モデル　Hierarchical Model



Ⓟ階層モデル

(モデル1) 単回帰



(モデル2) 階層モデル

学校ごとに切片が違う！
⇒ 勉強時間と成績の関係が強調！

(パート1)　$y_{ij} = \beta_{0j} + \beta_1 x_i + e_i$　← 重回帰

(パート2)　$\beta_{0j} = \gamma_0 + u_j$　← グループの独自性

u_j：学校ごとの独自性（○番の学校の効果）

γ_0：全ての学校の平均

📖 ノート３　統計量：偏回帰係数の解釈

（解釈のコツ１）係数の比較（標準偏回帰係数の利用）
（解釈のコツ２）点推定された偏回帰係数の意味

（１）解釈のコツ１：単位の影響を除く！

データ１ 非標準化データ

X1	X2	X3	Y
18.77	37	244	1.62
17.13	38	257	1.68
⋮	⋮	⋮	⋮
23.38	41	304	1.46

kg　km　g　kg

標準化 ⇒

データ２ 標準化データ

X1	X2	X3	Y
-1.21	-1.18	-1.67	1.62
-1.49	-1.12	-1.59	1.68
⋮	⋮	⋮	⋮
-0.42	-0.94	-1.29	1.46

① 非標準偏回帰係数（データ１を分析）
生データの単位をそのまま利用した回帰分析の結果。

	推定値	標準誤差	t 値	p 値	
$\hat{\beta}_0$	0.565	0.233	2.428	0.017	*
$\hat{\beta}_1$	0.044	0.012	3.781	0.000	***
$\hat{\beta}_2$	-0.004	0.004	-0.894	0.374	
$\hat{\beta}_3$	0.002	0.000	5.899	0.000	***

② 標準偏回帰係数（データ２を分析）
データを標準化したうえで行った回帰分析の結果。
※ 標準偏差の比が １ となり、相関係数のみに基づく。

	推定値	標準誤差	t 値	p 値	
$\hat{\beta}_0$	2.709	0.05	54.15	2e-16	***
$\hat{\beta}_1$	0.259	0.07	3.78	0.000	***
$\hat{\beta}_2$	-0.064	0.07	-0.89	0.374	
$\hat{\beta}_3$	0.385	0.07	5.90	5.23e-08	***

Ⓟ 偏回帰係数

$y_i = \underbrace{\beta_0}_{\text{切片}} + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$

傾き（偏回帰係数）

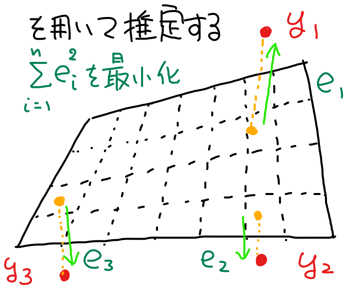
重回帰モデルの回帰係数を偏回帰係数と呼ぶ。

⇒単回帰の回帰係数と比べて扱いに少し注意が要る。

Ⓟ 偏回帰係数の推定

単回帰分析同様最小二乗法を用いて推定する

$\sum_{i=1}^{n} e_i^2$ を最小化



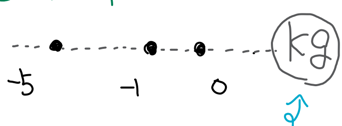
問 なぜ 0.002 の $\hat{\beta}_3$ は有意で 0.004 の $\hat{\beta}_2$ は有意にならないの？

$\underbrace{\hat{\beta}_3}_{0.002} \times \underbrace{x_{3i}}_{304} = 0.608$

$\underbrace{\hat{\beta}_2}_{-0.004} \times \underbrace{x_{2i}}_{41} = -0.164$

Ⓟ 標準化

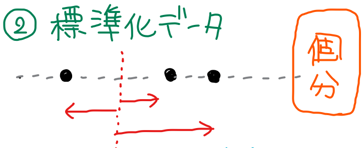
① 非標準化データ



もともとの単位に依存

⇓

② 標準化データ



平均から標準偏差何個分かで表わし直したデータ

Ⓟ論文を読むときのコツ

論文を読んでいて，

「『温度』の偏回帰係数は0.9で，有意だった。よって，『温度』は『背丈』に影響する」

のような記述に出会ったとき，

WAIT!

「結果に影響を与えそうな変数は，すべてモデルに含まれているのか？」

と問いながら冷静に判断.

（２）解釈のコツ２：偏回帰係数はモデル相対的！

① 解釈上の注意点

同じデータを用いても、ある条件を満たさない限り、重回帰の回帰係数は、単回帰の回帰係数とは一致しない

単回帰分析　$\underbrace{y_i}_{\text{背丈}} = \hat{\beta}_0 + \hat{\beta}_1 \underbrace{x_{1i}}_{\text{温度}} + e_i$

$\neq$　$\neq$

重回帰分析　$\underbrace{y_i}_{\text{背丈}} = \hat{\beta}_0 + \hat{\beta}_1 \underbrace{x_{1i}}_{\text{温度}} + \hat{\beta}_2 \underbrace{x_{2i}}_{\text{湿度}} + e_i$

② 理由

同じモデルに含まれているそのほかの要因（変数）に相対的に値が決まるため。

③ 解釈

例えば，「湿度」という影響を一定にして，考えなくてよくなったとき，「温度」が持つ効果

他の独立変数の影響が取り除かれた時、その独立変数が一単位増加すると従属変数がどの程度変化するのか。

④ 例外

独立変数同士に、関連性がない（無相関である）場合は、単回帰分析の結果と一致する。

※ただし、検定や信頼区間の結果は異なる。

例１：現地調査　研究者の人為的介入なし

ID	背丈 y	気温 x1	湿度 x2
1	10	20	40
2	11	18	50
3	11	19	50
4	12	22	65
5	13	26	60
6	15	28	70
7	14	27	80
8	20	35	85

独立変数の組みあわせは雑然!

① 単回帰分析　$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + e_i$

	推定値	標準誤差	t値	p値
beta0	0.181	1.564	0.116	0.912
beta1	0.536	0.062	8.554	0.000 ***

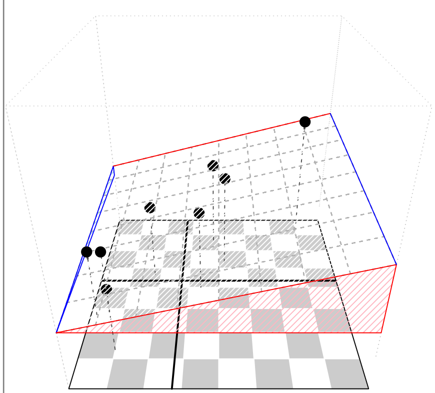
② 単回帰分析　$y_i = \hat{\beta}_0 + \hat{\beta}_2 x_{2i} + e_i$

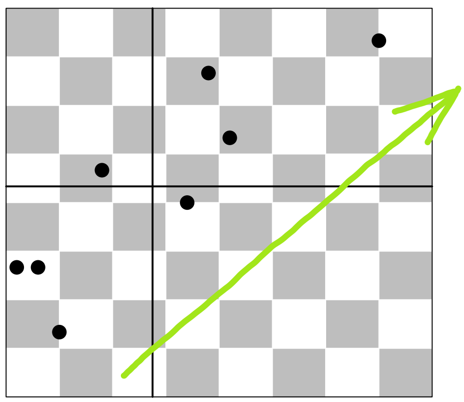
	推定値	標準誤差	t値	p値
beta0	2.037	2.603	0.782	0.464
beta2	0.179	0.041	4.423	0.005 **

③ 重回帰分析　$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + e_i$

	推定値	標準誤差	t値	p値
beta0	0.053	1.676	0.031	0.976
beta1	0.467	0.138	3.390	0.020 *
beta2	0.029	0.050	0.571	0.593

不一致





独立変数同士に相関あり!

例２：温室での調査　研究者が独立変数を調整可！

ID	背丈 y	気温 x1	湿度 x2
1	10	20	40
2	11	20	80
3	11	25	40
4	12	25	80
5	13	30	40
6	15	30	80
7	14	35	40
8	20	35	80

独立変数の組みあわせが整然

① 単回帰分析　$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + e_i$

	推定値	標準誤差	t値	p値
beta0	1.150	3.482	0.330	0.753
beta1	0.440	0.124	3.546	0.012 *

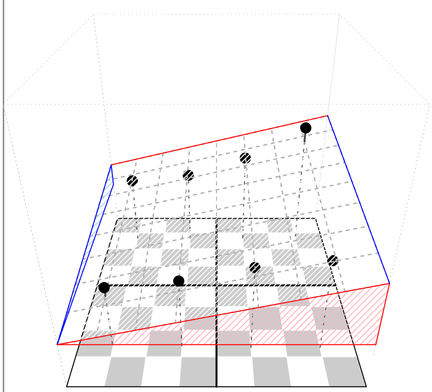
② 単回帰分析　$y_i = \hat{\beta}_0 + \hat{\beta}_2 x_{2i} + e_i$

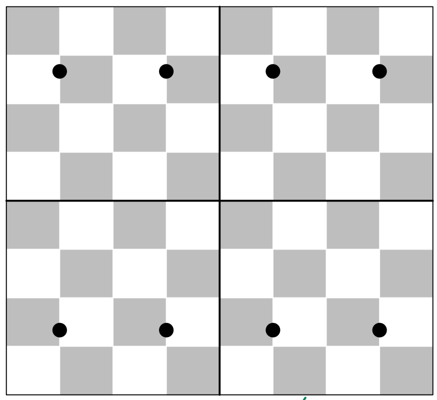
	推定値	標準誤差	t値	p値
beta0	9.500	3.506	2.710	0.035 *
beta2	0.063	0.055	1.127	0.303

一致

③ 重回帰分析　$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + e_i$

	推定値	標準誤差	t値	p値
beta0	-2.600	3.010	-0.864	0.427
beta1	0.440	0.092	4.778	0.005 **
beta2	0.063	0.026	2.428	0.060 .





独立変数たちが無相関

📖 ノート３　統計量：部分／偏相関係数と偏回帰係数

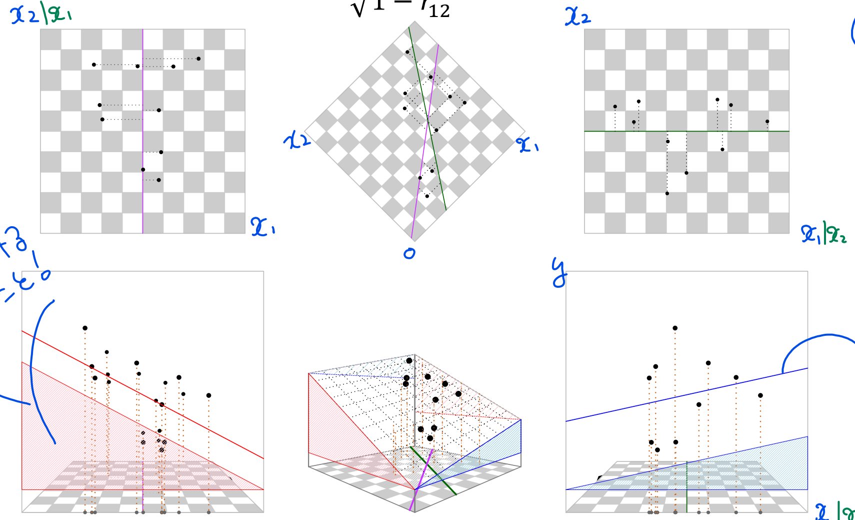
> ☞　（解釈のコツ１）点推定された偏回帰係数の意味
> （解釈のコツ２）係数の比較（標準偏回帰係数の利用）

（１）解釈のコツ１：それぞれの係数の意味を知る

① 部分相関係数 Part correlation coefficient

一つの変数から第三の変数の影響を除いた後で、二つの変数の相関を求めたもの。

$$r_{y(2|1)} = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{1 - r_{12}^2}}$$

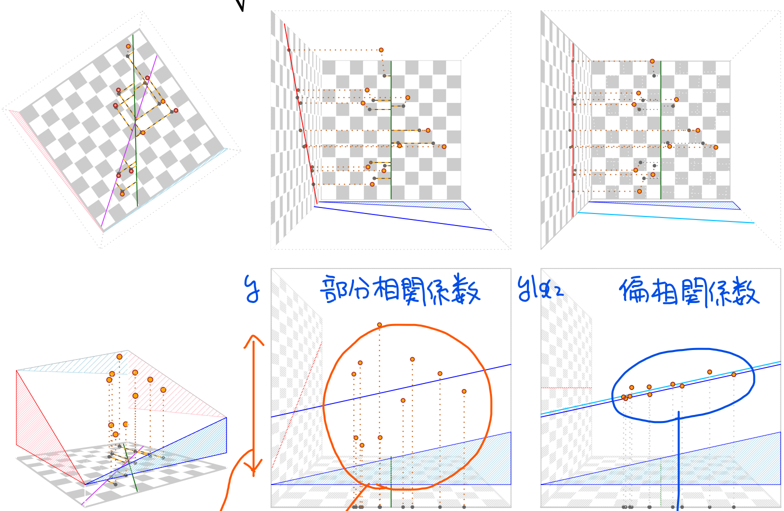


この散布図における相関係数のこと。

② 偏相関係数 Partial correlation coefficient

二つの変数それぞれから第三の変数の影響を除いた後で、それら二つの変数の相関を求めたもの。

$$r_{y2|1} = \frac{1}{\sqrt{1 - r_{y1}^2}} \times \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{1 - r_{12}^2}}$$



S_y　$r_{y(1|2)}$　$S_{1|2}$

この相関を表わした

Ⓟ 偏回帰係数

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$

切片　傾き

（偏回帰係数）

重回帰モデルの回帰係数を偏回帰係数と呼ぶ。

⇒単回帰の回帰係数と比べて扱いに注意が必要!

Ⓟ 偏回帰係数

「他の独立変数の影響をとりのぞいたときに」

その独立変数がもつ効果の大きさを表わす。

Ⓟ 記法：$x_1 | x_2$　バー

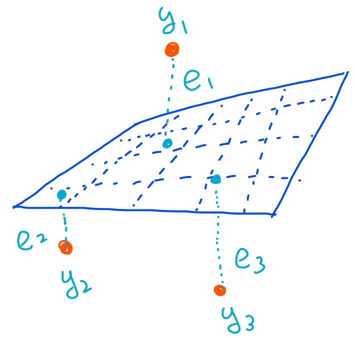
「x_2の影響が取り除かれたx_1」を

$x_1 | x_2$

とりのぞかれる変数

より純粋なものになる。

P 重回帰における最小乗法



$\sum_{i=1}^{n} e_i^2$ を最小化

P 単回帰の係数

$\hat{\beta}_1 = r_{yx} \dfrac{s_y}{s_x}$

P 記法：$r_{y(1|2)}$

$r_{y(x_1|x_2)}$

$= r_{y(1|2)}$

単にわかりやすさを優先して、"x"を省略して表現しただけ。

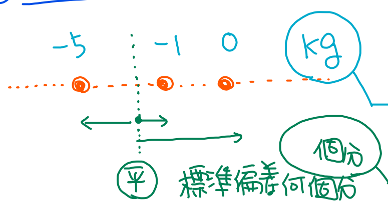
問 なぜ、0.002の$\hat{\beta}_3$は有意で、-0.004の$\hat{\beta}_2$は有意ではないの？

⇒データの値の大きさがまちまち。

$\underbrace{\hat{\beta}_3}_{0.002} \times \underbrace{x_{3i}}_{304} = 0.608$

$\underbrace{\hat{\beta}_2}_{-0.004} \times \underbrace{x_{2i}}_{41} = -0.164$

P 標準化



もともとの単位に依存するデータ（非標準化データ）

平均から標準偏差何個分かに表わし直した（標準化データ）

③ **偏回帰係数** Partial regression coefficient

ある独立変数からそれ以外の独立変数の影響を除いた残差変数によって従属変数を予測するときの回帰係数。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

└ x_1の影響をとりのぞいたx_2がyに対してもつ影響

点推定値

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2$$

$$\hat{\beta}_1 = r_{y(1|2)} \times \frac{s_y}{s_{1|2}} \qquad s_{1|2} = s_1\sqrt{1 - r_{21}^2}$$

$$\hat{\beta}_2 = r_{y(2|1)} \times \frac{s_y}{s_{2|1}} \qquad s_{2|1} = s_2\sqrt{1 - r_{12}^2}$$

※ 偏回帰係数と部分／偏相関係数の関係

縦軸と横軸の(i) 相関係数$r_{y(1|2)}$と(ii)標準偏差の比$s_y/s_{1|2}$から構成されている（参照：第4講）

（2）**解釈のコツ２**：単位の影響を除く！

データ１ → 標準化 → データ２（標準化データ）

X1	X2	X3	Y
18.77	37	244	1.62
17.13	38	257	1.68
⋮	⋮	⋮	⋮
23.38	41	304	1.46

X1	X2	X3	Y
-1.21	-1.18	-1.67	1.62
-1.49	-1.12	-1.59	1.68
⋮	⋮	⋮	⋮
-0.42	-0.94	-1.29	1.46

① 非標準偏回帰係数（データ１を分析）

生データの単位をそのまま利用した回帰分析の結果。

	推定値	標準誤差	t 値	p 値	
$\hat{\beta}_0$	0.565	0.233	2.428	0.017	*
$\hat{\beta}_1$	0.044	0.012	3.781	0.000	***
$\hat{\beta}_2$	-0.004	0.004	-0.894	0.374	
$\hat{\beta}_3$	0.002	0.000	5.899	0.000	***

② **標準偏回帰係数**（データ２を分析）

データを標準化したうえで行った回帰分析の結果。

※ 標準偏差の比が 1 となり、相関係数のみに基づく。

	推定値	標準誤差	t 値	p 値	
$\hat{\beta}_0$	2.709	0.05	54.15	2e-16	***
$\hat{\beta}_1$	0.259	0.07	3.78	0.000	***
$\hat{\beta}_2$	-0.064	0.07	-0.89	0.374	
$\hat{\beta}_3$	0.385	0.07	5.90	5.23e-08	***

📖 ノート４　分析の評価１：作った統計モデルの正確さを評価

（参照：第４講ノート５）

データへの適合度

独立変数を複数取り入れたことで、従属変数の予測はどれくらい正確になったの？

（１）平方和／分散の直交分解

予測値$\hat{y}$と残差eは互いに無相関であるので、その和である$y=\hat{y}+e$の分散を次のように分解することができる。

① 分散の直交分解

$$s_y^2 = s_{\hat{y}}^2 + s_e^2$$

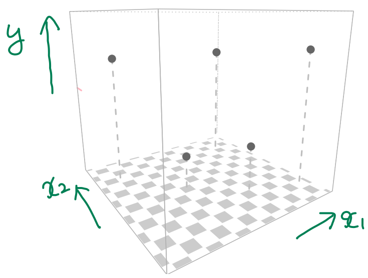
② 平方和の直交分解

$$\frac{1}{n}ss_y = \frac{1}{n}ss_{\hat{y}} + \frac{1}{n}ss_e$$

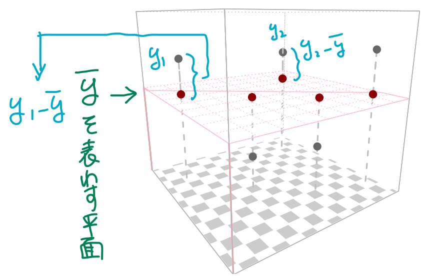
両辺を $\frac{1}{n}$ で割る

$$\underbrace{ss_y}_{\text{観測値の平方和}} = \underbrace{ss_{\hat{y}}}_{\text{予測値の平方和}} + \underbrace{ss_e}_{\text{残差の平方和}}$$

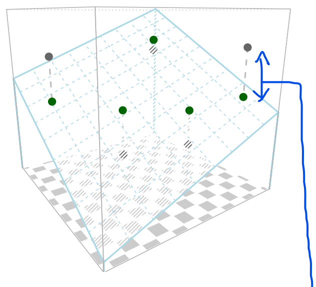
A 観測値の姿(分布)



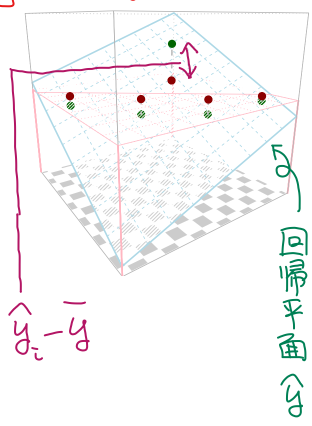
B 観測値yの分散



C 残差eの分散



D 予測値$\hat{y}$の分散



Ⓟ復習：分散の分解

$$S_{A+B}^2 = S_A^2 + \boxed{2S_{AB}} + S_B^2$$

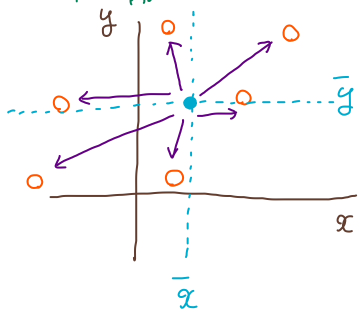
ここが0。無相関のとき

$$S_{A+B}^2 = S_A^2 + S_B^2$$

y　$\hat{y}$　e

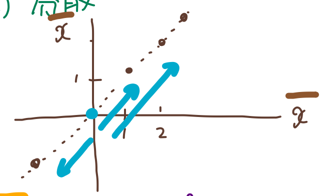
Ⓟ復習：平方和と分散

(1) 共分散



$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

(2) 分散



$$\boxed{\frac{1}{n}}\sum_{i=1}^{n}(x_i-\bar{x})^2 = S_x^2$$

(3) 平方和

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = SS_x$$

分散と平方和の差

$$S_x^2 = \frac{1}{n}SS_x$$

Ⓟ平方和の分解のイミ

これは、

B ＝ C ＋ D

を表わしている。

Ⓟ重相関係数の考え方

(ケース1) 予測値と観測値が近い

予測が上手くいっている!

$r_{y\hat{y}}$ = 高い!

(ケース2) 予測値と観測値が遠い

予測がうまくいってない!

= 回帰直線が データにフィットしてない!

$r_{y\hat{y}}$ = 低い

⇒そこで、「データと回帰直線がどのくらいフィットしてるか」という指標に $r_{y\hat{y}}$ を用いる。

Ⓟ復習：決定係数

① 分散の直交分解

$$S_y^2 = S_{\hat{y}}^2 + S_e^2$$

こいつが S_y^2 においてどのくらいの割合を示めるの?

② 決定係数

$$1 = \frac{S_{\hat{y}}^2}{S_y^2} + \frac{S_e^2}{S_y^2}$$

$$\frac{S_{\hat{y}}^2}{S_y^2} = 1 - \frac{S_e^2}{S_y^2}$$

Ⓟモデル比較

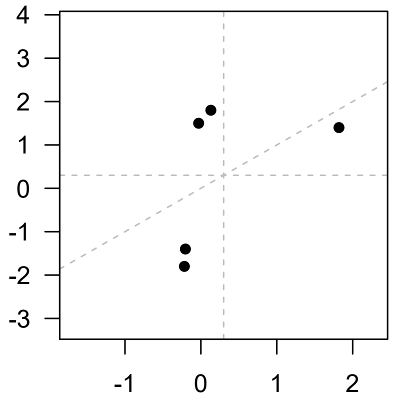
決定係数でモデル比較はできない!

（2） **重相関係数 R** Multiple correlation coefficient

これは、実測値 y と 予測値 $\hat{y}$ の相関係数であり、R と表す。

$$R = r_{y\hat{y}}$$

二乗する → 決定係数 R^2



（3） **決定係数** Coefficient of determination

これは独立変数がどれだけ従属変数の値を決定するかを示す指標。別名：分散説明率 Proportion of variance accounted for

$$R^2 = 1 - \frac{s_e^2}{s_y^2}$$

（4） **決定係数の性質**

独立変数を増やすと追加した独立変数が従属変数の予測に寄与しているかに関係なく決定係数は単調に増加する。

Model 1	$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$	: $R^2_{モデル1}$
Model 2	$y_i = \beta_0 + \beta_1 x_{1i} \boxed{+ \beta_2 x_{2i}} + \varepsilon_i$	: $R^2_{モデル2}$
Model 3	$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \boxed{+ \beta_3 x_{3i}} + \varepsilon_i$	: $R^2_{モデル3}$

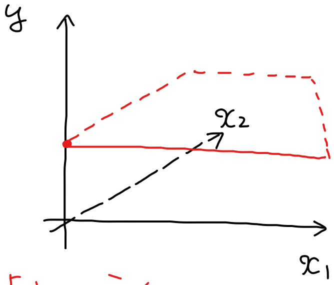
$R^2_{モデル1} \le R^2_{モデル2} \le R^3_{モデル3}$

→ 決定係数は増加!

📖 ノート５ 分析の評価２：係数の推定値の正確さの評価

Ⓟ 重回帰式「全体」の検定

(帰無仮説)



「すべての方向において傾きは 0 (=水平だ)」

(対立仮説)

「すべての方向において傾きが0だ」という訳ではない

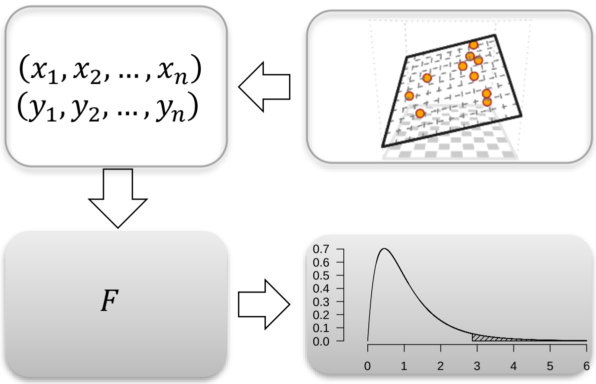
⇔ 傾きが0ではない方向が存在する

ステップ１　重回帰式「全体」の検定
データから何らかの統計量を計算し、その標本分布を用いて「$\beta_1 = \beta_2 = \cdots = \beta_p = 0$（係数が全部 0）」という帰無仮説を検証したい。

ステップ２　「個別」の回帰係数の検定
データから何らかの統計量を計算し、その標本分布を用いて「i 番目の偏回帰係数について$\beta_i = 0$だ」という帰無仮説を検証したい。

ステップ３　重回帰式・偏回帰係数の区間推定
重回帰式やそれぞれの偏回帰係数について、信頼区間や予測区間を計算し、幅を持った推定を行う。

（1）検定１：重回帰式「全体」の検定　ステップ１



① 帰無仮説と対立仮説

H0：「$\beta_1 = \beta_2 = \cdots = \beta_p = 0$（すべての係数が０）」
H1：「$\beta_1, \beta_2, \ldots, \beta_p$の少なくとも一つはゼロではない」

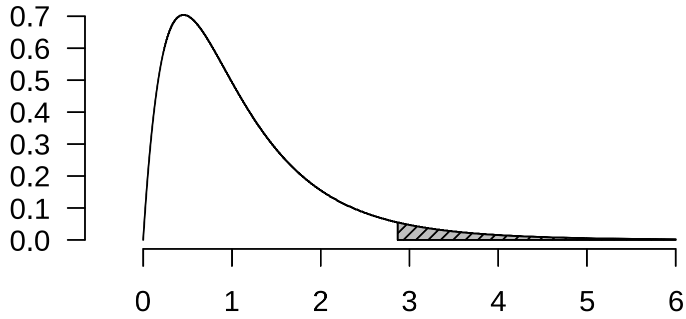
② 検定統計量Ｆ（重回帰式の検定における検定統計量）

$$F=\frac{ss_{\hat{y}}/p \qquad \text{「}\hat{y}_{\mathrm{i}}\text{の全体平均からのばらつき」}}{ss_e/(n-p-1) \quad \text{「}\mathrm{y}_{\mathrm{i}}\text{の}\hat{y}_{\mathrm{i}}\text{からのばらつき」}}$$

（※pは独立変数の数）

③ Ｆ分布

H0 が真なら F 値は自由度$p, n-p-1$の F 分布に従う。



当然出てくるであろう疑問

疑問１：F 分布ってどこから登場したの？

⇒ $N(0,1)$からカイ二乗分布を経由して登場

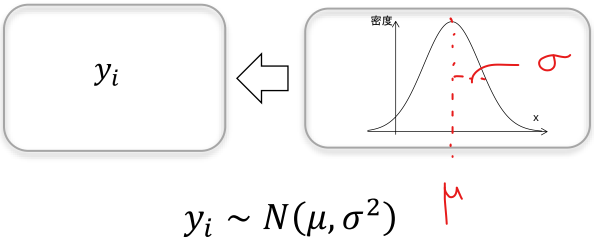
疑問２：自由度ってなにものなの？

⇒ カイ二乗分布の再生性に起因して登場

※ カイ二乗分布とＦ分布

① 正規分布

これは、ランダムな誤差が積み重なって登場する分布。



$$y_i \sim N(\mu, \sigma^2)$$

② 標準正規分布

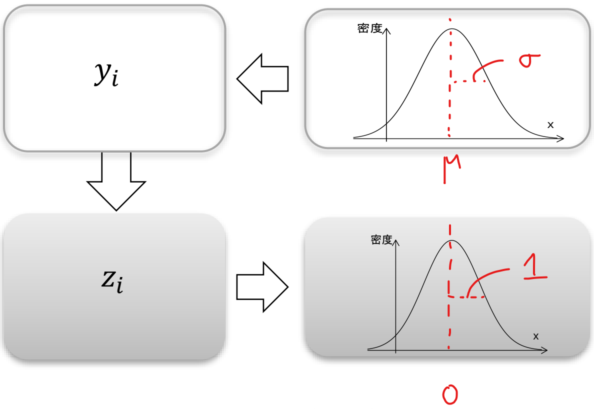
これは、平均が 0、分散が 1 の正規分布$N(0,1)$。正規分布$N(\mu, \sigma^2)$に従う変数y_iを標準化したz_iが従う分布。

Ⓟ 標準化

$y_i \sim N(\mu, \sigma^2)$

$\boxed{\dfrac{y_i - \mu}{\sigma}} \sim N(0,1)$

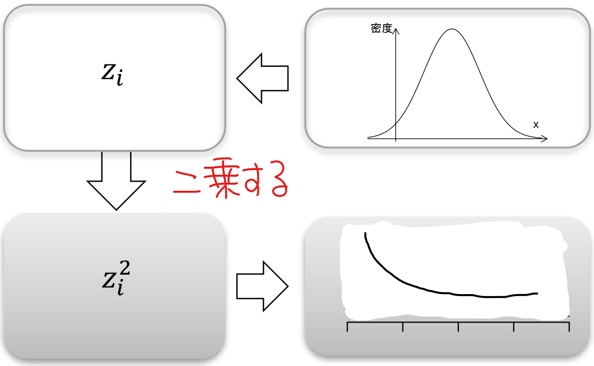
$= z_i$



$$z_i \sim N(0, 1)$$

③ カイ二乗分布

これは$N(0,1)$に従う変数を二乗した統計量が従う分布。



$$z_i^2 \sim \chi^2(1)$$

自由度1のχ²分布.

Ⓟ 自由度

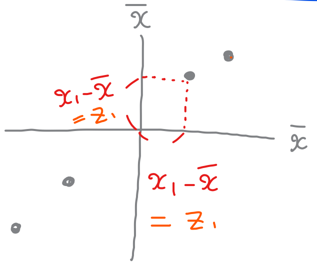
これは，何個 χ^2 分布にしたがう変数を足しあわせたのかを表わす。

$$z_1^2 \sim \chi^2(1)$$

$$z_1^2 + z_2^2 \sim \chi^2(2)$$

$$\vdots$$

Ⓟ $\chi^2(n)$ を考える訳



$$\frac{1}{n}\left(z_1^2 + z_2^2 + z_3^2 + z_4^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} z_i^2$$

$$= \text{分散}$$

再生性（正規分布）

$y_1 \sim N(\mu_1, \sigma_1^2)$ 英語の点

$y_2 \sim N(\mu_2, \sigma_2^2)$ 数学の点

$\vdots$

+) $y_n \sim N(\mu_n, \sigma_n^2)$ 古文の点

$y_1 + y_2 + \cdots + y_n$ 総合得点

$\sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$

再生性（カイ二乗分布）

$$z_1^2 \sim \chi^2(1)$$

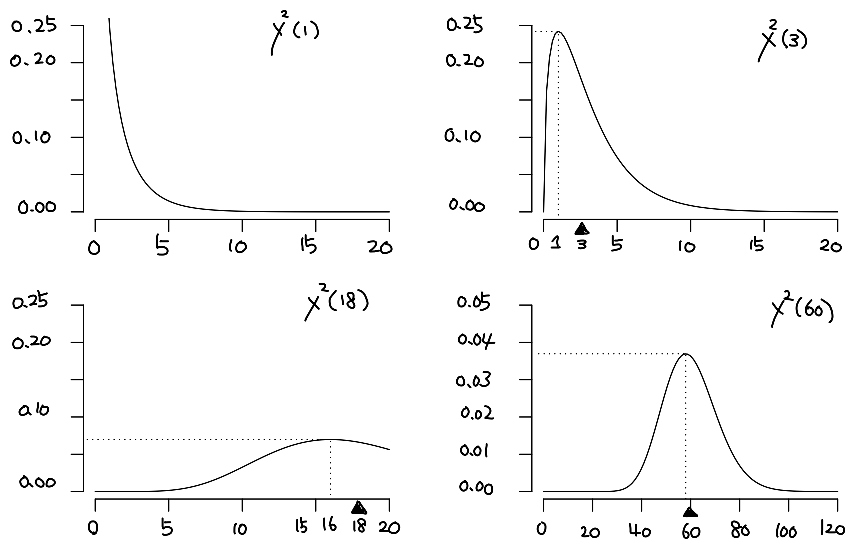
$$z_2^2 \sim \chi^2(1)$$

$$\vdots$$

+) $z_n^2 \sim \chi^2(1)$

$$z_1^2 + z_2^2 + \cdots + z_n^2 \sim \chi^2(n)$$

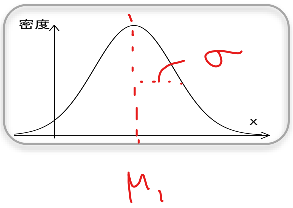
カイ二乗分布



④ F 分布

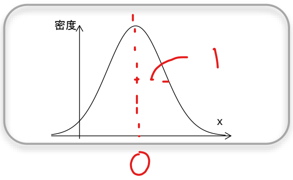
これは、互いに独立な、二つのカイ二乗分布に従う変数 $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$の比が従う分布。

$$y_{1i}$$



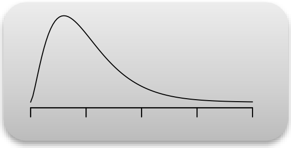
$N(\mu_1, \sigma_1^2)$

$$z_{1i} = \frac{y_{1i} - \mu_1}{\sigma_1}$$



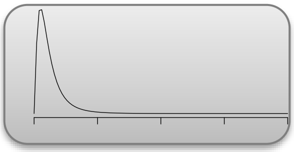
$N(0, 1)$

$$\chi_1^2 = z_{11}^2 + z_{12}^2 + \cdots + z_{1n_1}^2$$



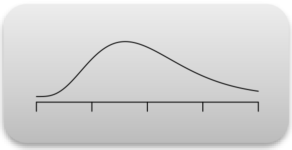
$\chi^2(n_1)$

$$F = \frac{\frac{1}{n_1}\sum_{i=1}^{n_1} z_{1i}^2}{\frac{1}{n_2}\sum_{i=1}^{n_2} z_{2i}^2}$$



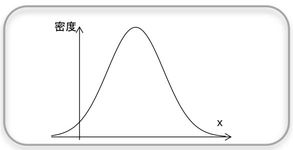
$F(n_1, n_2)$

$$\chi_2^2 = z_{21}^2 + z_{22}^2 + \cdots + z_{2n_2}^2$$



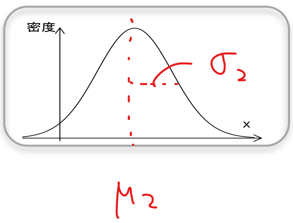
$\chi^2(n_2)$

$$z_{2i} = \frac{y_{2i} - \mu_2}{\sigma_2}$$



$N(0, 1)$

$$y_{2i}$$



$N(\mu_2, \sigma_2^2)$

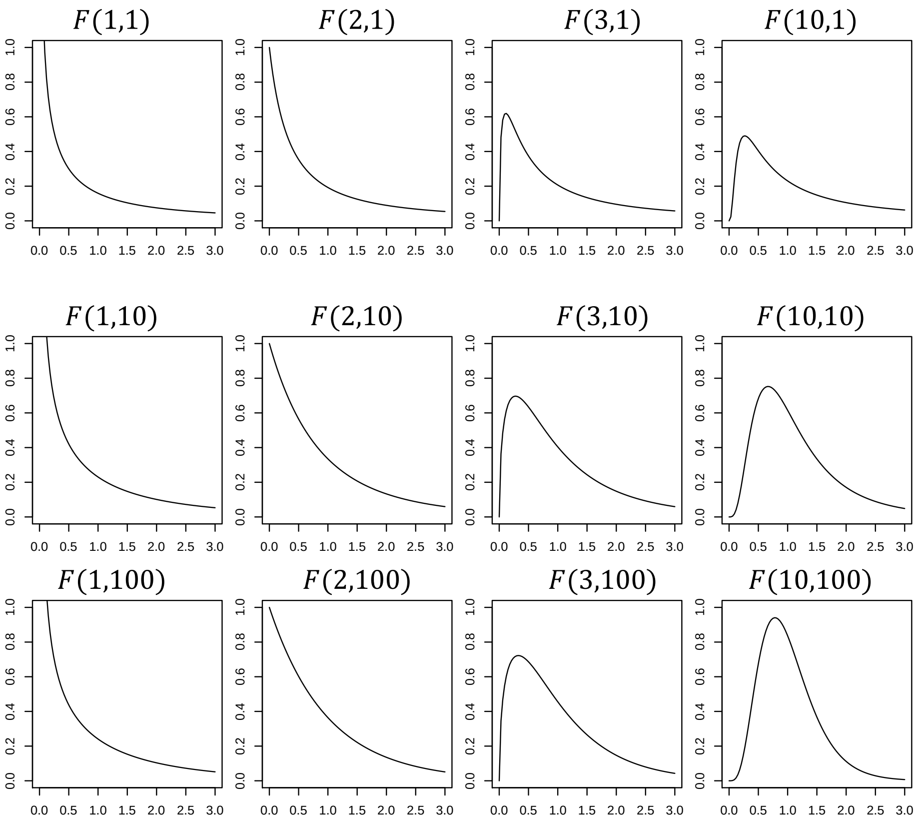
定義：F 値

$$F = \frac{\frac{1}{n_1}\chi_1^2}{\frac{1}{n_2}\chi_2^2} = \frac{\frac{1}{n_1}\sum_{i=1}^{n_1} z_{1i}^2}{\frac{1}{n_2}\sum_{j=1}^{n_2} z_{2j}^2}$$

$$= \frac{\frac{1}{n_1}\left(z_{1i}\text{の平方和}\right)}{\frac{1}{n_2}\left(z_{2j}\text{の平方和}\right)}$$

$$= \frac{z_{1i}\text{の分散}}{z_{2j}\text{の分散}}$$

F 分布



A 標準化する際には中心を確認！（統計量y_iを標準化）

$$y_i \sim N(\mu, \sigma^2)$$

人間の視点

$$\frac{y_i - \bar{y}}{\sigma} \nsim N(0, 1)$$

$$\sum_{i=1}^{n} \left(\frac{y_i - \bar{y}}{\sigma} \right)^2 \nsim \chi^2(n)$$

全知全能の視点

$$\frac{y_i - \mu}{\sigma} \sim N(0, 1)$$

$$\sum_{i=1}^{n} \left(\frac{y_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

B 標準化する際には中心を確認！（統計量y_iを変換）

$$y_i \sim N(\mu, \sigma^2)$$

$$\sum_{i=1}^{n} \left(\frac{y_i - \bar{y}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

C 標準化する際には中心を確認！（統計量 $\bar{y}$を標準化）

$$y_i \sim N(\mu, \sigma^2)$$

中心極限定理

$$\bar{y} \sim N\left(\mu, \frac{\sigma^2}{n} \right)$$

$$\frac{\bar{y} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

$$\left(\frac{\bar{y} - \mu}{\sigma / \sqrt{n}} \right)^2 \sim \chi^2(1)$$

$\boxed{\text{B}}$ において自由度が n-1 になる理由

$$\underbrace{\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2}_{\boxed{\text{A}}\ \uparrow\ \chi^2(n)} = \underbrace{\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{\sigma}\right)^2}_{\boxed{\text{B}}\ \downarrow\ \chi^2(n-1)} + \underbrace{\left(\frac{\bar{y}-\mu}{\sigma/\sqrt{n}}\right)^2}_{\boxed{\text{C}}\ \uparrow\ \chi^2(1)}$$

ということは
再生性から

$\chi^2(n-1)$

証明

$$\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}\left\{\frac{(y_i-\bar{y})+(\bar{y}-\mu)}{\sigma}\right\}^2$$

$$=\sum_{i=1}^{n}\left\{\left(\frac{y_i-\bar{y}}{\sigma}\right)^2+\frac{2(y_i-\bar{y})(\bar{y}-\mu)}{\sigma^2}+\left(\frac{\bar{y}-\mu}{\sigma}\right)^2\right\}$$

$$=\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{\sigma}\right)^2+\frac{2(\bar{y}-\mu)}{\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y})+\sum_{i=1}^{n}\left(\frac{\bar{y}-\mu}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{\sigma}\right)^2+\frac{2(\bar{y}-\mu)}{\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y})+n\left(\frac{\bar{y}-\mu}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{\sigma}\right)^2+\frac{2(\bar{y}-\mu)}{\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y})+\left(\frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$

$$=\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{\sigma}\right)^2+\left(\frac{\bar{y}-\mu}{\sigma/\sqrt{n}}\right)^2$$

（２）検定２：「個別」の回帰係数の検定 ステップ２

① 帰無仮説と対立仮説

H0：$\beta_1 = 0$

H1：$\beta_1 \neq 0$

② 検定統計量ｔ

☞ ｔ値（回帰係数の検定における検定統計量）

$$t = \frac{\hat{\beta}_i \text{「}\beta_i\text{の推定値」}}{\hat{\sigma}_{\hat{\beta}_i} \text{「}\beta_i\text{の推定値」の標準誤差の推定値}}$$

	点推定値	標準誤差	t value	Pr(>|t|)	
$\hat{\beta}_0$	-0.49	0.08	-6.30	1.13e-06	***
$\hat{\beta}_1$	1.02	0.15	6.66	4.63e-07	***

F-statistic: 36.57 on 3 and 98 DF, p-value: 6.032e-16

（3）偏回帰係数の信頼区間

標本を繰り返し抽出した時、この範囲を設けておけば 100 回に 95 回、真のβ_0, β_1を含むだろう、という区間。

$$y_i = \boxed{\beta_0} + \boxed{\beta_1} x_{1i} + \boxed{\beta_2} x_{2i} + \cdots + \varepsilon_i$$

それぞれの偏回帰係数の信頼区間

（4）予測値の 95%信頼区間　図の灰色の区間

標本を繰り返し抽出した時、点x_jに対してこの範囲を設けておけば 100 回に 95 回、μ_jを含むだろう、という区間。

$$y_i = \boxed{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots} + \varepsilon_i$$

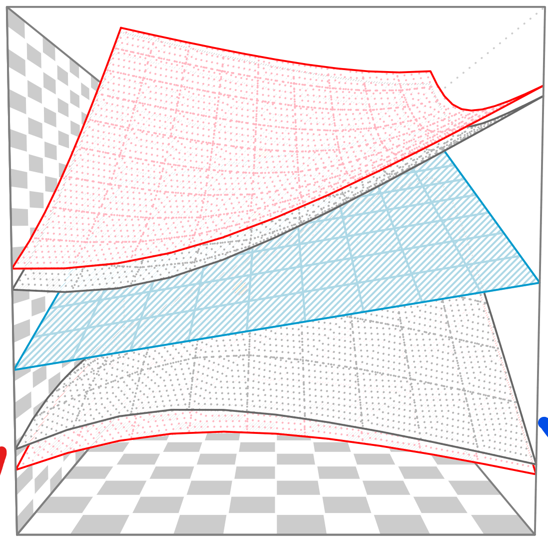
$\hat{y}$の信頼区間

（5）データの 95%予測区間　図の赤色の区間

標本を繰り返し抽出した時、点x_jに対してこの範囲を設けておけば 100 回に 95 回データを含むだろう、という区間。

$$\boxed{y_i} = \boxed{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \varepsilon_i}$$

データyの予測区間



yという（将来の）観測値の予測区間

$\hat{y}$という回帰平面の信頼区間

📖 ノート６ 分析の評価３：想定した統計モデルの適切さの評価
（モデルの改良）

例：入試当日の得点の予想

P個の独立変数（要因）

- x_1 全国模試の点数
- x_2 共通テストの点数
- x_3 12 月の勉強時間
- ⋮
- x_p 試験会場への遠さ

→ 入試総合得点 y

（１）**目的**：素性選択 Feature selection
重要な独立変数（素性）のみに絞りたい！

（動機１） **過学習** overfitting
独立変数が多いと訓練データへの依存度が強まり新標本への対応力が落ちてしまう。

（動機２） **解釈可能性** interpretability
独立変数が多いと解釈が難しくなる。

自然言語処理

工学的興味：予測精度の向上

トレード・オフの関係にある。

理論的興味：有意味な解釈の追及

理論言語学

（動機３） **多重共線性** Multicollinearity
独立変数が多いと独立変数が互いに強い関連性を持つ非理想的な状況が生まれやすい。

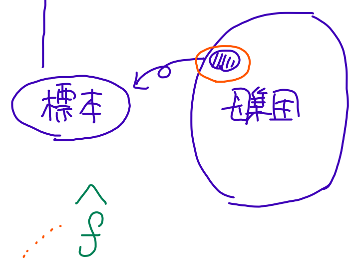
⇒ 偏回帰係数の推定量が不安定になる。

多数存在する独立変数を適切に操作しより良いモデルを模索するための方法として次の三つの方略がよく紹介される。

案１ （モデル縮約）： 重要度の低い変数を削除する
案２ （モデル比較）： よりスマートなモデルを探す
案３ （合成指標） ： 変数同士を合成し新指標を作る

Ⓟ 統計学における「学習」
研究者が想定したモデルが
研究者が採取したデータを
学習

Ⓟ 過学習 overfitting
研究者が想定したモデルが
研究者が採取したデータを
学習しすぎること



独立変数の数が多いと、
左（標本）にとてもとてもフィットするが、右（母集団）から
離れてしまう危険性がある！

Ⓟ 多重共線性の特徴

どちらか一方のみが大きい値をとるので、(小さい)

$\hat{\beta}_1$ 大　$\hat{\beta}_2$ 小

$\hat{\beta}_1$ 小　$\hat{\beta}_2$ 大

のどちらでもいい。

$\hat{\beta}_1$は小さかったり大きかったりするので、推定が不安定！

（２）多重共線性　Multicollinearity

独立変数が互いに強い関連性を持ち、偏回帰係数の推定量が不安定になってしまう状態のこと。

ある独立変数の値が大きくなると、別の独立変数の値が大(小)に予測される

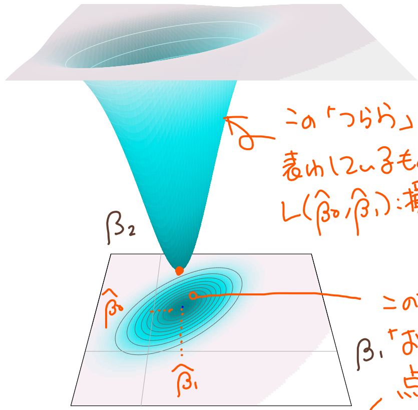
① 独立変数同士の相関が高い場合

例：店舗の売り上げの予測

x_1　月間の雨天の日数 → y　店舗の売り上げ

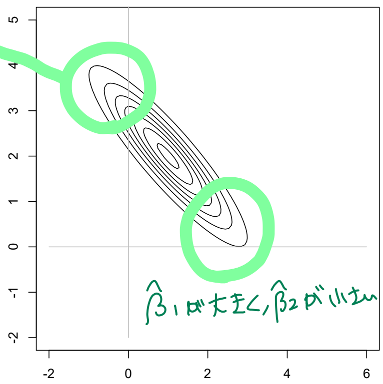
x_2　月間の降雨量 → y　店舗の売り上げ

② 点推定の幾何的な理解

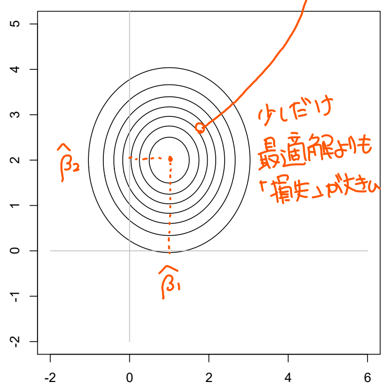


$\hat{\beta}_1$が小さく、$\hat{\beta}_2$が大きい

（ケース１）多重共線性



（ケース２）



Ⓟ VIFの発想

(ステップ1) j番目以外の独立変数でj番目の独立変数を予測！

x_j ← $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p$

x_jが他の変数と強い関連ありの場合、この重回帰式

$$x_j = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_{j-1}x_{j-1} + \alpha_{j+1}x_{j+1} + \cdots + \alpha_p x_p$$

の決定係数 $R_{x_j|x_{j-1}}$ は大

(ステップ2) 分数(比の値)をつくっています。

③ 分散拡大要因 VIF（Variation Inflation Factor）

多重共線性の診断に使われる指標。10 以上のものは危険だが、4 程度の数値でも多重共線性に注意するといい。

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R^2_{x_1|x_2}}$$

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R^2_{x_j|-x_j}}$$

大　小　大

（2）モデル縮約 Model Shrinkage（正則化 Regularization）

P 正則化項

① 基本的なアイデア

モデルの複雑さを反映する罰則（正則化項）を設けた上で推定を行い、偏回帰係数を最小化・削除を行う。

独立変数の数の多さのこと。

(例1) ラッソ回帰

$\hat{\beta}_1 + \hat{\beta}_2 \le 10$

3.8　10.1

$\hat{\beta}_1 + \hat{\beta}_2 \le 10$

2.9　7.1

(例2) リッジ回帰

$\hat{\beta}_1^2 + \hat{\beta}_2^2 \le 10$

通常の最小二乗推定

残差の平方和

$$L(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^{n} e_i^2$$

損失関数

を最小化

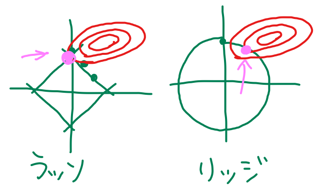
正則化項付きの最小二乗推定

$R(\beta_0, \beta_1, \dots, \beta_p) \le s$という制約を満たす範囲で

正則化項

$$L(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^{n} e_i^2$$

を最小化



② ラッソ回帰 Lasso Regression

$R(\beta_0, \beta_1, \dots, \beta_p)$の部分にL1 ノルムを用いたもの。

単純な総和

P L1 ノルム

$R(\beta_0, \beta_1, \dots, \beta_p) = \sum_{j=1}^{p} |\beta_j|$

$\sum_{i=1}^{n} \beta_i \le s$という制約を満たす範囲で

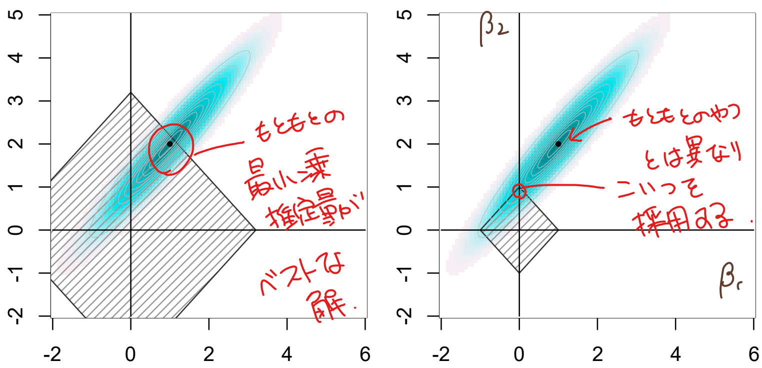
$$L(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^{n} e_i^2$$

を最小化

λを大きくすると正則化項の罰則が増大。β_jたちは大きい値を取りづらくなり効果の小さいものから 0 になる。

P L2 ノルム

$R(\beta_0, \beta_1, \dots, \beta_p) = \sum_{j=1}^{p} (\beta_j)^2$



P ひし形の頂点が解のとき

$(\hat{\beta}_1, \hat{\beta}_2) = (0, 1)$

のようになったときは、

$\hat{\beta}_1$は、モデルの中で効果なし

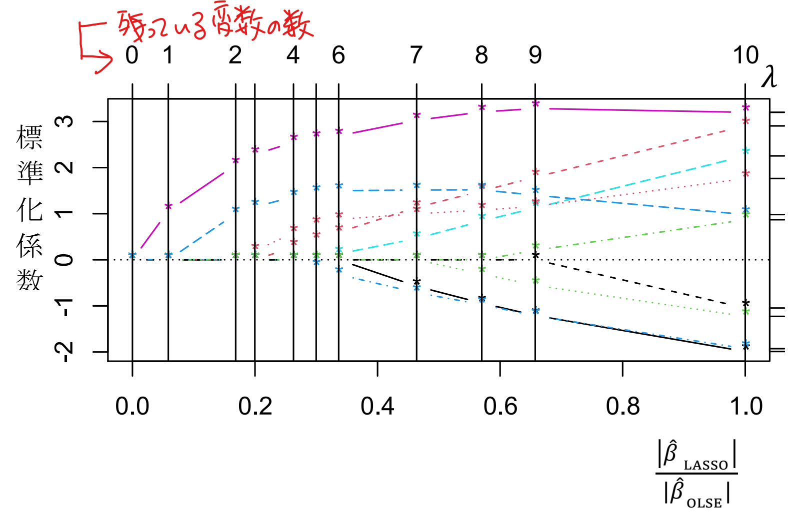
と判断されたことになる。

⇒ 独立変数の数を減らす

③ 実践上の手続き

A. チューニング

λの値によって、通常の最小二乗推定量（OLSE）と比べてどのくらい係数の値が小さくなるのかを吟味する。



※係数の比較：標準化回帰係数にしておくのがコツ。

B. 変数の数の選択　　　（☞ クロスバリデーション）

クロスバリデーションを行い、最もパフォーマンスが高いモデルを採用することが多い。

```
> library(glmnet)
> data(QuickStartExample)

> fit <- glmnet(x, y)
> plot(fit)
> coef(fit, s = 0.1)

> cvfit <- cv.glmnet(x, y)
> plot(cvfit)
> coef(cvfit, s = "lambda.min")
> predict(cvfit, newx = x[1:5,], s = "lambda.min")
```

https://glmnet.stanford.edu/articles/glmnet.html

（3）モデル選択 Model selection

これは、考えられるモデルたちの中で、最もよいパフォーマンスを持つモデルを探す試みのこと。

モデル選択のステップ

手順１：モデルを比較するための基準を決める
手順２：考えられるモデルたちのパフォーマンスを測る
手順３：最良のパフォーマンスを持つモデルを解釈する

① モデルのパフォーマンスを測る指標

（系統１）訓練データの誤差を修正した指標

[1] 赤池情報量基準（AIC） 小

$$AIC = \frac{1}{n\hat{\sigma}^2}(RSS + 2d\hat{\sigma}^2)$$

[2] ベイズ情報量基準（BIC） 小

$$BIC = \frac{1}{n}(RSS + \log(n)\, d\hat{\sigma}^2)$$

[3] Mallow’s C_p 小

$$C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2)$$

[4] 自由度調整済み決定係数 大

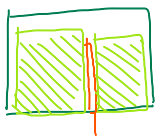
$$R^2_{adjusted} = 1 - \frac{RSS/(n-d-1)}{TSS/(n-1)}$$

（系統２）テストデータの誤差を計算した指標

[1] Leave-One-Out Cross-Validation 小

(= LOOCV)

$$CV_{(n)} = \frac{1}{n}\sum_{j=1}^{n} MSE_j^2$$

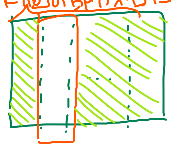


1つだけをテストデータ.

[2] K-fold CV 小

K個の部分集合にわける.

$$CV_{(k)} = \frac{1}{k}\sum_{j=1}^{k} MSE_j^2$$



Ⓟ「考えられる」モデルとは?

今研究者がp個の独立変数を用意した.

$M_1 : \beta_0$

⋮

$M_{(2^p-1)} : \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1}$

$M_{2^p} : \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$

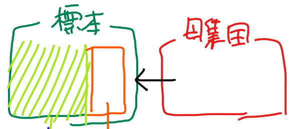
1 2 3 … p-1 p

Y/N Y/N Y/N … Y/N Y/N

$2 \times 2 \times 2 \times \cdots \times 2$

$= 2^p$ 個のモデルを作れる

Ⓟ 訓練データとテストデータ



推定に用いるデータ

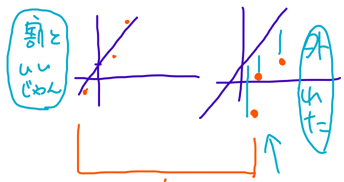
(訓練データ)

$\hat{f}$ (訓練データを用いて作ったモデル)

$\hat{f} = \hat{y}_i = 1 + 2x_i$

テストデータ

残しておいたデータ



Ⓟ pが大きいとき

$2^{20} = 1{,}048{,}576$個

⇒コンピュータへの負荷を考えて,全2を見ることはあきらめ,効率的に良いモデルを探したい.

$1 + \frac{20 \times 21}{2} = 211$

② ベストなモデルを見つけるためのアルゴリズム

（アルゴリズム１）Best Subset Selection
2^p個のモデルすべてを比較する

（アルゴリズム２）Stepwise Selection
モデルの部分集合を作り勝ち抜き試合式で$1 + p(p+1)/2$個を比較

ステップ1

総当たり法

	x_1	x_2	x_3	x_4	
M1	×	×	×	×	0
M2	○	×	×	×	1
M3	×	○	×	×	1
M4	×	×	○	×	1
M5	×	×	×	○	1
M6	○	○	×	×	2
M7	○	×	○	×	2
M8	○	×	×	○	2
M9	×	○	○	×	2
M10	×	○	×	○	2
M11	×	×	○	○	2
M12	○	○	○	×	3
M13	○	○	×	○	3
M14	○	×	○	○	3
M15	×	○	○	○	3
M16	○	○	○	○	4

ステップ3

M4

（前のページの基準でBESTを決める）

ステップ2

変数増加法

		x_1	x_2	x_3	x_4
M1	★	×	×	×	×
M2		○	×	×	×
M3		×	○	×	×
M4	★	×	×	○	×
M5		×	×	×	○
M7		○	×	○	×
M9		×	○	○	×
M11	★	×	×	○	○
M14	★	○	×	○	○
M15		×	○	○	○
M16	★	○	○	○	○

このグループの中ではM4がBest

このグループの中ではM11がBest

このグループではM14がBest

③ 実践上の手続き

```
> library(leaps)
> longley = data.frame(scale(longley))
> y = longley$Employed; x = longley[,1:6]
> r1 = regsubsets(x,y)
> l1 = leaps(x,y,nbest =1,method ="r2")
> plot(r1); plot(l1$size, l2$adjr2, type = "b")
```