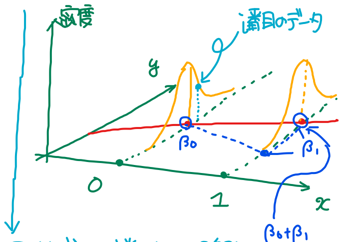
Ⓟ ギリシア文字 β と ε
b → β（ベータ）
e → ε（エプシロン）

📖 ノート１　母集団に対する仮定：統計モデル

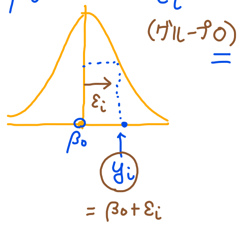
Ⓟ 統計モデルの把握

①図（幾何的な理解）

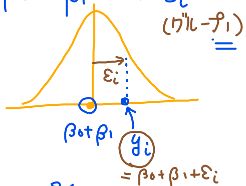


②数式（代数的な理解）

ⓐ $y_i = \beta_0 + \varepsilon_i$
(グループ0)
=



ⓑ $y_i = \beta_0 + \beta_1 + \varepsilon_i$
(グループ1)
=



この ⓐ と ⓑ を場合わけせずに表現することを考える。

ⓒ $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$x_i = \begin{cases} 0 : \text{グループ}0 \\ 1 : \text{グループ}1 \end{cases}$

（１）**統計モデル**

これは、研究者が母集団に対して想定するモデル。

例１：t 検定

① **ダミー変数**

分類を表す質的変数（名義尺度変数）に便宜的に 0、1, …のように数値を与えコード化した変数。

女性　A
日本人　1

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is } \_\_\_\_ \\ 0 & \text{if } i\text{th person is not } \_\_\_\_ \end{cases}$$

女性　A
日本人　1

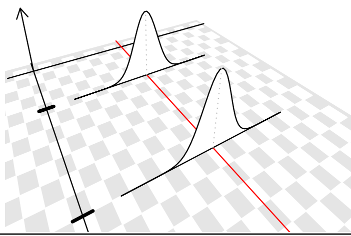
※1 か 0 のものを**指示変数 Indicator variable** と呼ぶ。

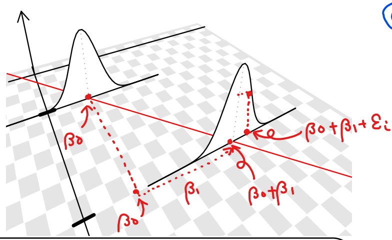
② 線形モデル

ダミー変数

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$= \begin{cases} \beta_0 + \beta_1 + \varepsilon_i & \text{(グループ1)} \\ \beta_0 \qquad\quad + \varepsilon_i & \text{(グループ2)} \end{cases}$$





質問　ダミー変数はいつも０と１なんですか？

この他の組み合わせも可能です。ただ、係数の解釈が変わります。例えば、次の例ではβ_0は全体の平均、β_1は平均から上下に男女がどのくらい離れるかを表します。

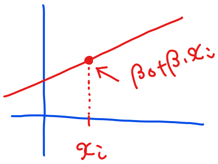
$$x_i = \begin{cases} 1 & \text{if } i\text{th person is } \_\_\_\_ \\ -1 & \text{if } i\text{th person is not } \_\_\_\_ \end{cases}$$

$$y_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 - \beta_1 + \varepsilon_i \end{cases}$$

Ⓟ 誤差項 Error Term

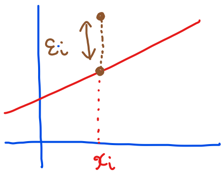
① 赤い直線

これは，母集団の構造的な傾向



② 茶色の矢印

これは赤い直線から「外れた」誤差．



Ⓟ 正規分布

これは，誤差がつみかさなると生まれる分布．

⇒ 誤差項が従う分布として仮定されている！

Ⓟ 分布を表わすやり方

(i) $y_i = \beta_0+\beta_1 x_i + \varepsilon_i$

$\varepsilon_i \sim N(0, \sigma^2)$

(ii) $y_i \sim N(\mu_i, \sigma^2)$

$\mu_i = \beta_0+\beta_1 x_i$

(iii) $y_i \sim N(\beta_0+\beta_1 x_i, \sigma^2)$

例２：単回帰分析

① 概要

これは「予測変数（x）と応答変数（y）が一次直線と基本とする関係にある」と想定する統計モデル。

② 線形モデル

注 $y_i \sim N(\underbrace{\beta_0+\beta_1 x_i}_{x_i\text{の値に依存}}, \sigma^2)$

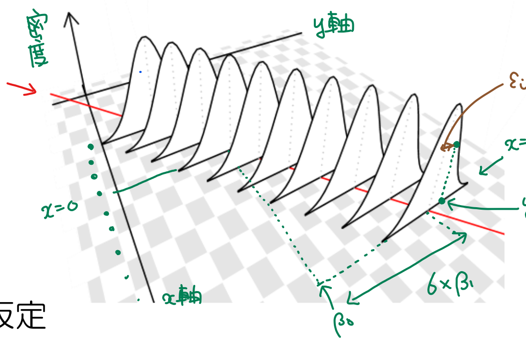
$$y_i = \beta_0 + \beta_1 \underbrace{x_i}_{i\text{番目のデータの}x\text{の値}} + \underbrace{\varepsilon_i}_{\text{誤差項}}$$

$\varepsilon_i \sim N(0, \sigma^2)$

①分散：添え字 i に依存しない！

②平均：常にゼロ！

（cf. σ_i^2 と書いたら i ごとに別の値をとるということだが，今回は，i にかかわらず σ^2．）



$y_i = \beta_0+\beta_1 \underbrace{x_i}_{x_i=6} + \underbrace{\varepsilon_i}$

山の頂上から i 番目のデータが どのくらい離れているか という，i 番目のデータの独自性を表す．

（２）仮定

【標本の抽出の仕方に関する仮定】

① **独立性の仮定**：　各要素は互いに独立

② **同一分布の仮定**：　標本は同じ分布から抽出

③ **無作為性の仮定**：　標本はランダムに抽出

$$y_i \overset{i.i.d}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$$

【母集団の姿に関する仮定】

① 正規性の仮定：　母集団の分布は正規分布

② 分散の等質性の仮定：　~~[illegible]~~母集団の分散は同じ

x_i の値に関係なく．

📖 ノート２　統計量：回帰係数の点推定（最小二乗法）

（１）最小二乗法 Least Squares Method

┌ 平方ユークリッド距離（平均で使っていた距離）

これは、 二乗基準 で測ったデータと直線の適合の悪さ（残差）を最小化することで点推定値を出す方法。

$\hat{\beta}_0$と$\hat{\beta}_1$の値

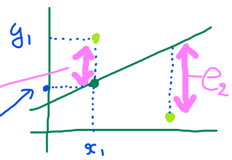
距離（遠さ）.

① 残差 Residuals

これは、予測の誤差のこと。

┌ i番目の観測値の残差

$y_1-(\hat{\beta}_0+\hat{\beta}_1 x_1)= e_1$



$\hat{\beta}_0+\hat{\beta}_1 x_1 = \hat{y}_i$

$$e_i = y_i - \hat{y}_i$$

y_iのデータに対する予測値

$$= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

i番目のyの観測値　人間が提案回帰直線から作った予測値

② 損失関数 Loss function

これは、適合の悪さを測る関数。一般的な単回帰では残差平方和 Residual Sum of Squares を損失関数とする。

$$L(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} e_i^2 = (e_1^2+e_2^2+e_3^2+\cdots+e_n^2)$$

この二つの組 $\hat{\beta}_0$と$\hat{\beta}_1$をとっかえひっかえしたものの中で、最も距離が小さいのを選ぶ

$$= \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

（２）最小二乗推定量

これは、最小二乗法で求めた点推定量のこと。単回帰モデルの場合は、次のような結果になる。

① 傾き

$$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2} = \frac{xとyの共分散}{xの分散}$$

$$= r_{xy}\frac{s_y}{s_x} = xとyの相関係数 \times \frac{yの標準偏差}{xの標準偏差}$$

② 切片

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

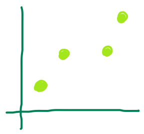
③ 誤差

$$\hat{\sigma}^2 = \frac{1}{n-2} \times \sum_{i=1}^{n} e_i^2$$

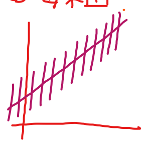
$$= \frac{1}{n-2} \times \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

Ⓟ 母集団の直線と標本の直線

① 標本



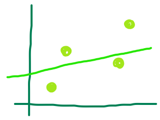
② 母集団



$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$\varepsilon_i \sim N(0, \sigma^2)$

③ 統計量



$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$

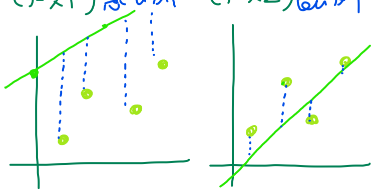
赤：人間には見えない値

緑：人間が計算して、提案・推測する値

限られたデータしか見えない人間が、どうやって、赤い直線の候補となる、緑色の直線（$\hat{\beta}_0, \hat{\beta}_1$の値）を提案するの？

Ⓟ 最小二乗法の基本的なアイデア

（ケース1）悪い例　（ケース2）良い例



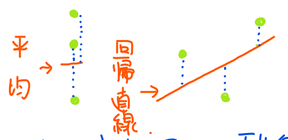
「良い悪い」を判断する基準として、"損失"として、各点から直線までの「距離」を考える。

平方ユークリッド距離（二乗距離）

そして、「距離」が一番短い直線を選ぼう！、と考えるのが、「最小二乗法」である。

二乗基準で測った距離を最小化する方法.

Ⓟ 平均と回帰直線の関係



どちらも点線の長さを二乗基準で最小化

Ⓟ 回帰係数 $\hat{\beta}_1$

$$\hat{\beta}_1 = r_{xy} \times \frac{S_y}{S_x}$$

(i) 相関(高) → $\hat{\beta}_1$(高)

$$\underbrace{\hat{\beta}_1}_{\nearrow} = \underbrace{r_{xy}}_{\nearrow} \times \frac{S_y}{S_x}$$

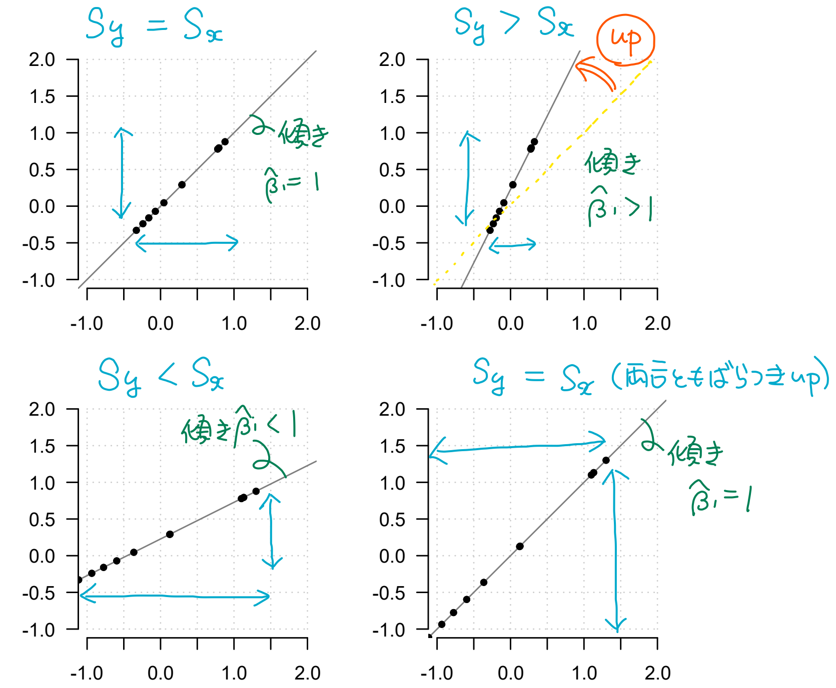
(ii) $S_y > S_x$ → $\hat{\beta}_1$(高)

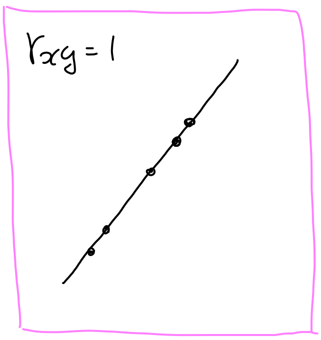
yの分散がxの分散より大きいなら，傾きも大.

$$\underbrace{\hat{\beta}_1}_{\nearrow} = r_{xy} \times \frac{S_y}{\underbrace{S_x}_{\nearrow}}$$

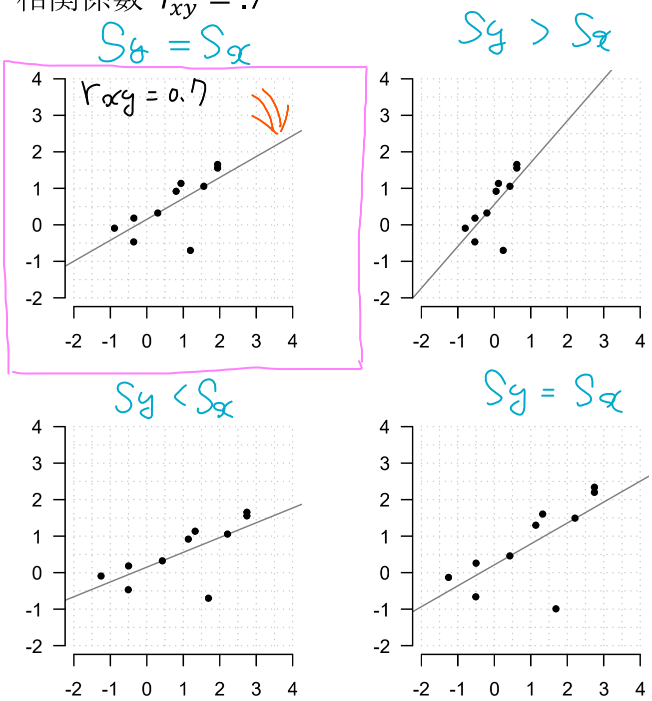
（３）回帰係数と相関係数の関係

① 相関係数 $r_{xy} = 1$





② 相関係数 $r_{xy} = .7$



📖 ノート３　統計量：点推定と推定量の性質

（１）点推定

① 点推定

母集団に想定した統計モデルのパラメータの値がどのような値なのか点（統計量一つ）で推定する統計的推測。

② 推定量 Estimator

これは、点推定に採用される統計量のこと。

③ 推定量の作り方

これには、様々な方法が考案されている。

（例１）モーメント法

★（例２）最小二乗法

（例３）最尤推定法

× 小さい標本の特性

○ くり返り標本をとったときの特性

（２）推定量の選択基準Ⅰ：小標本特性

これは、標本を何度も何度も繰り返しとった場合に生じるメリットのこと（サンプルサイズは小さくてもよい）。

「ヒストグラムを精密なものとした時」

① 不偏性 unbiasedness

これは、その推定量の期待値が母パラメータに一致すること。不偏性を持つ推定量を不偏推定量という。

㊟ $\bar{x} \longrightarrow \mu$ の不偏推定量

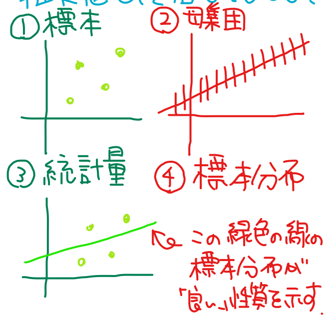
$\hat{\beta}_0 \longrightarrow \beta_0$ 〃

$\hat{\beta}_1 \longrightarrow \beta_1$ 〃

② 有効性 Efficient estimator

これは、不偏性を持つ推定量の中で分散が最小のもののこと。有効性を持つ推定量を有効推定量という。

(問) なぜ最小二乗法で求めた $\hat{\beta}_0$ と $\hat{\beta}_1$ の値を，母集団の推定値として信じていいの？

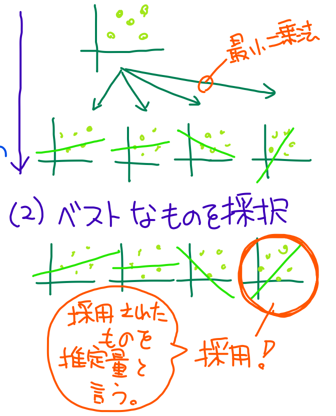


(答) 最小二乗推定量には，「不偏性」を中心とする望ましい性質があるから．

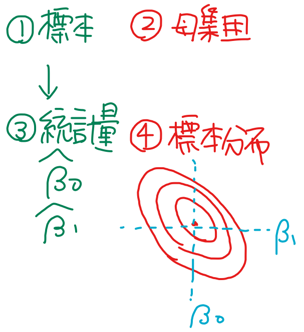
(P) 点推定の2つのステップ

(1) 標本から多数統計量を作る

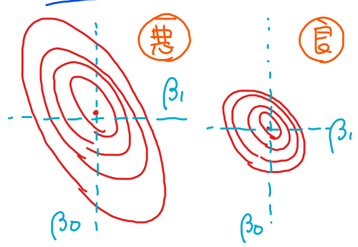
(2) ベストなものを採択



(P) 不偏推定量



(P) 有効性



（３）推定量の選択基準Ⅱ：大標本特性
これは、サンプルサイズ n の大きい標本を採ったときに生まれるメリットのこと。

① 漸近的不偏性 asymptotic unbiasedness
これは、サンプルサイズが大きいとき、その推定量の期待値が母パラメータに一致すること。
㊟ 標本分散は、「漸近的不偏性」はもつが、「不偏性」はもたない!

② 一致性
これは、サンプルサイズが大きいとき、推定量と母パラメータに差が生じる確率は極めて小さくなること。

③ 漸近的有効性 asymptotic efficiency
これは、サンプルサイズが大きいときに、不偏推定量の中で最小分散となること。

（４）分散と不偏分散

① 分散
これは、漸近的不偏性を持つが不偏性は持たない。

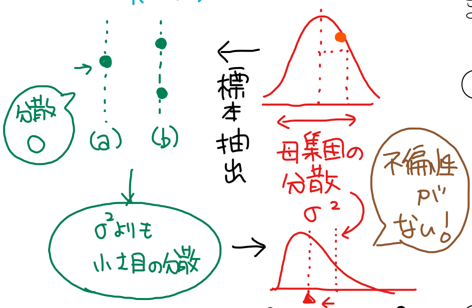
$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \hat{\sigma}_1^2$$

② 不偏分散　Unbiased variance
これは、不偏性を持ち、当然漸近的不偏性も持つ。

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \hat{\sigma}_2^2$$

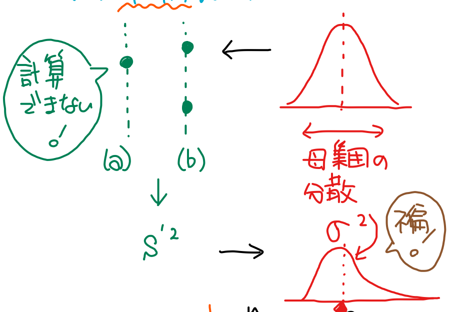
Ⓟ 分散の種類

(1) (標本)分散



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

(2) 不偏分散



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

㊟ ばらつきを生じさせているのは、1つ目のdata pointではなく、2つ目のデータ(=(b))だと言える。

Ⓟ バイアス Bias

$E[\hat{\sigma}^2]$: $\hat{\sigma}^2$の期待値 (重心).

① $E[\hat{\sigma}_1^2]-\sigma^2 \neq 0$ ← バイアス有!

② $E[\hat{\sigma}_2^2]-\sigma^2 = 0$ ← バイアスなし!

📖 ノート４　分析の評価１：係数の推定値の正確さの評価

☞　分析を評価する三つの軸

（評価軸１）　係数の推定値の正確さを評価

（評価軸２）　作った統計モデルの正確さを評価

（評価軸３）　想定した統計モデルの適切さを評価

Ⓟ 回帰分析

(問1) HOW MUCH？
$\hat{\beta}_0, \hat{\beta}_1$ はどのくらい？
→ 最小二乗法 (1-1-2)
→ 理由：不偏性 (1-1-3)

(問2) WHETHER？
$\hat{\beta}_1$ は, 0 に等しいと見なして良いのか？ (1-1-4)

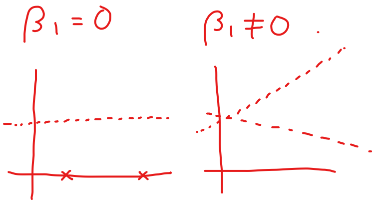
Ⓟ 仮説検定のすすめ方

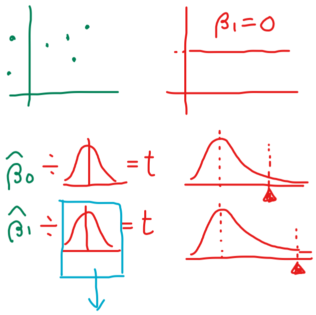
（１）回帰係数の検定

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）

$$t = \frac{\hat{\beta}_1 \quad \text{「}\beta_1\text{の推定値」}}{\hat{\sigma}_{\hat{\beta}_1} \quad \text{「}\beta_1\text{の推定値」の標準誤差の推定値}}$$

$\hat{\beta}_1$ の したがう 標本分布！
こいつの 標準誤差 が，

$\sigma_{\hat{\beta}_1}$

さらに, これを, 人間が 推定したものを,

$\hat{\sigma}_{\hat{\beta}_1}$ と表わす.

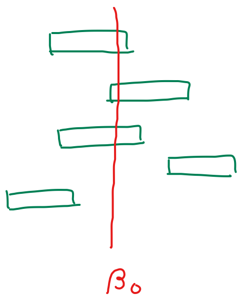
$$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{1}{n-2} \times \frac{ss_e}{ss_x}}$$

	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	***

（２）回帰係数の信頼区間

Ⓟ 信頼区間も統計量

$$\left[\ \hat{\beta}_1 - t_c \times \hat{\sigma}_{\hat{\beta}_1},\ \ \hat{\beta}_1 - t_c \times \hat{\sigma}_{\hat{\beta}_1}\ \right]$$



100回に 95回は
β0 という真の値を含む
区間を作った！

	点推定値	下限	上限
$\hat{\beta}_0$	-0.49	-0.65	-0.33
$\hat{\beta}_1$	1.02	0.70	1.33

Ⓟ 三つのタイプの区間

① β たちの信頼区間

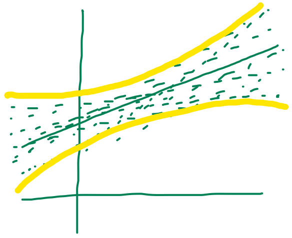
$y_i = \boxed{\hat{\beta}_0} + \boxed{\hat{\beta}_1} x_i + e_i$

−0.65 〜 −0.33　　0.70 〜 1.33

② $\hat{y}$ の 信頼区間

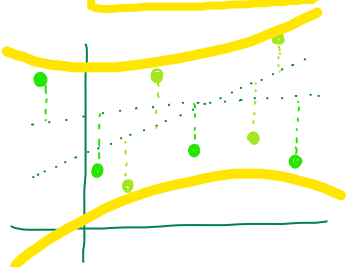
$y_i = \boxed{\hat{\beta}_0 + \hat{\beta}_1 x_i} + e_i$

$\hat{y}_i$



③ y の 予測区間

$y_i = \boxed{\hat{\beta}_0 + \hat{\beta}_1 x_i + e_i}$



（3）予測値の 95%信頼区間　**図の破線の区間**

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

$$\left[\ \hat{y}_p - t_c \times \hat{\sigma}_{\hat{y}_p},\ \ \hat{y}_p - t_c \times \hat{\sigma}_{\hat{y}_p}\ \right]$$

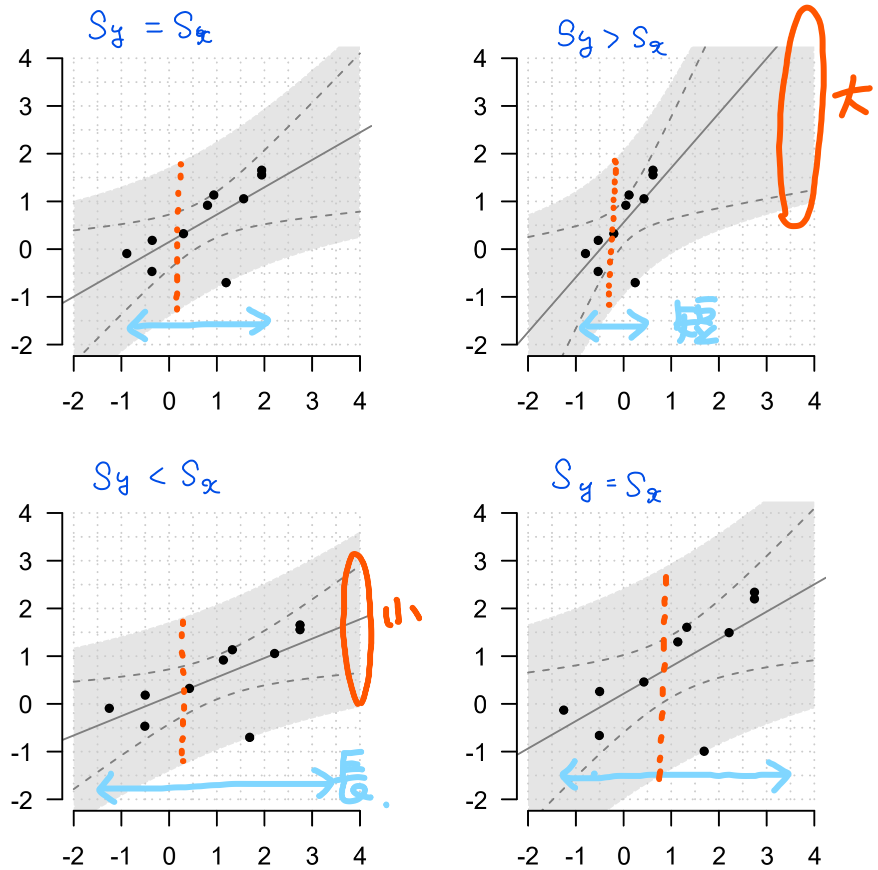
$$\hat{\sigma}_{\hat{y}_p} = \sqrt{\frac{ss_e}{n-2} \times \left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{ss_x}\right]}$$

（4）データの予測区間　**図の灰色の区間**

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

$$\left[\ y_p - t_c \times \hat{\sigma}_{y_p},\ \ y_p - t_c \times \hat{\sigma}_{y_p}\ \right]$$

$$\hat{\sigma}_{y_p} = \sqrt{\frac{ss_e}{n-2} \times \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{ss_x}\right]}$$



注 右のデータは、「ノート2」の $r_{xy} = 0.7$ というデータについて、予測区間を信頼区間を求めたもの。

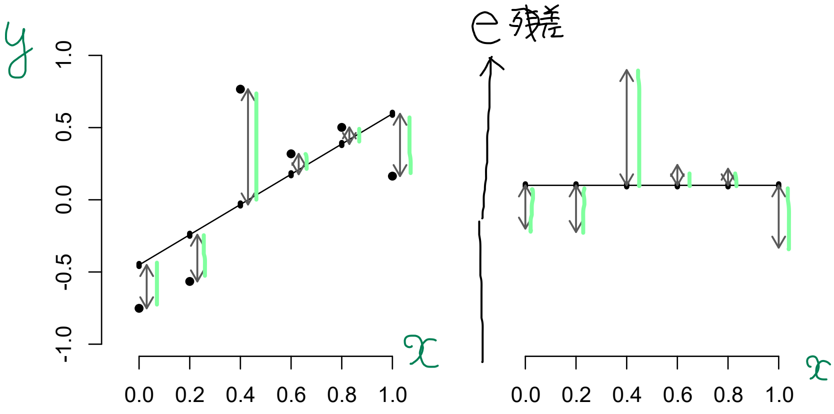
より安定した推論を行いたいなら、xに関しては幅広く取る!

作りあげた直線がどのくらい
データにフィットしているのか考える。

📖 ノート５ 分析の評価２：作った統計モデルの正確さを評価

（１）残差プロット

これは、x_iの値と残差e_iの関係をプロットした散布図。



（２）残差の性質

① 独立変数と残差の相関

独立変数xと残差eの相関r_{xe}は、必ずゼロになる。

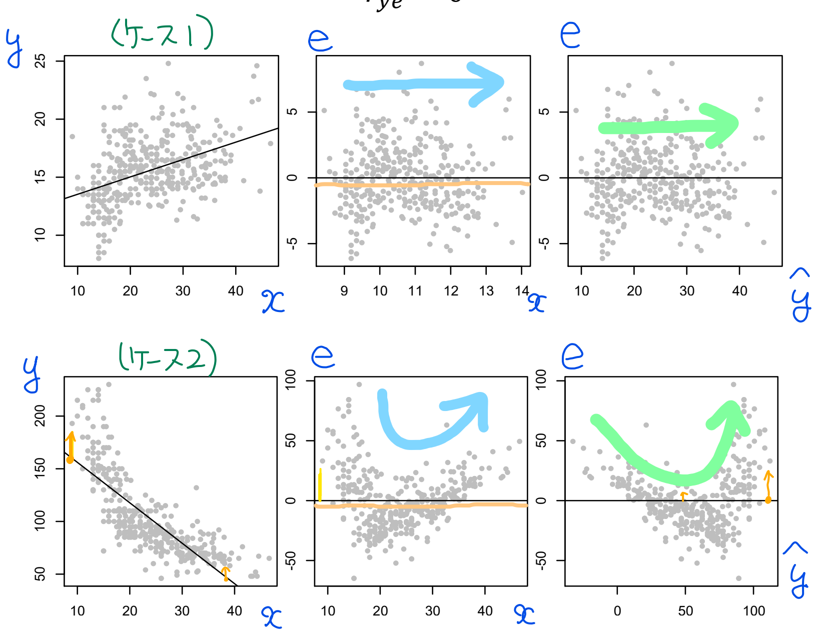
$$r_{xe} = 0$$

← 右肩上がりでも！
右肩下がりでもない！

② 予測値と残差の相関

予測値$\hat{y}$と残差eの相関$r_{\hat{y}e}$は、必ずゼロになる。

$$r_{\hat{y}e} = 0$$



残差プロット

（３）変数の直交分解

① 分散の分解（一般的な場合）

$$s_{a+b}^2 = s_a^2 + 2s_{ab} + s_b^2$$

$(a+b)^2 = a^2 + 2ab + b^2$ のような関係がある!

② 分散の分解（a と b が無相関の時）

$$s_{a+b}^2 = s_a^2 + 0 + s_b^2$$

なくなる!

③ 従属変数yの分散の分解

無相関!

$y = \hat{y} + e$

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

データの分散　予測値の分散　残差の分散

（４）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

= 直線（研究者が推定したモデル）が，どのくらいデータにフィットしているのかを表わす指標

考え方

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

↓ 両辺を S_y^2 で割る.

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

↓ 移項する.

$$\frac{S_{\hat{y}}^2}{S_y^2} = 1 - \frac{S_e^2}{S_y^2}$$ =（決定係数）.

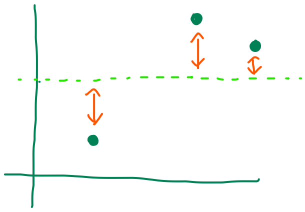
決定係数 $R^2 = 1 - \frac{s_e^2}{s_y^2}$

この値が大きいほど，作った直線は，データにフィットしていると言える.

$0 \leq R^2 \leq 1$

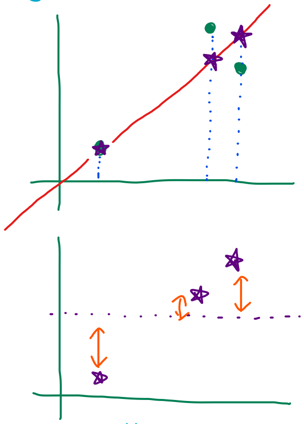
P yの分散の分解

① yの分散

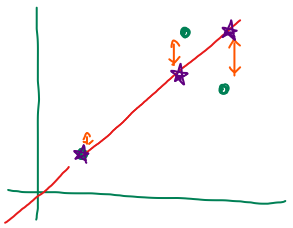


これは，回帰直線を引こうが引くまいが データをとったら決まる量

② $\hat{y}$の分散



③ eの分散



注 決定係数と相関係数

$r_{y\hat{y}}$: yと$\hat{y}$の相関係数

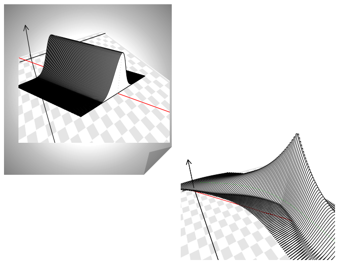
$(r_{y\hat{y}})^2 = R^2$

📖 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

（回帰診断法）

母集団に対する仮定の検証

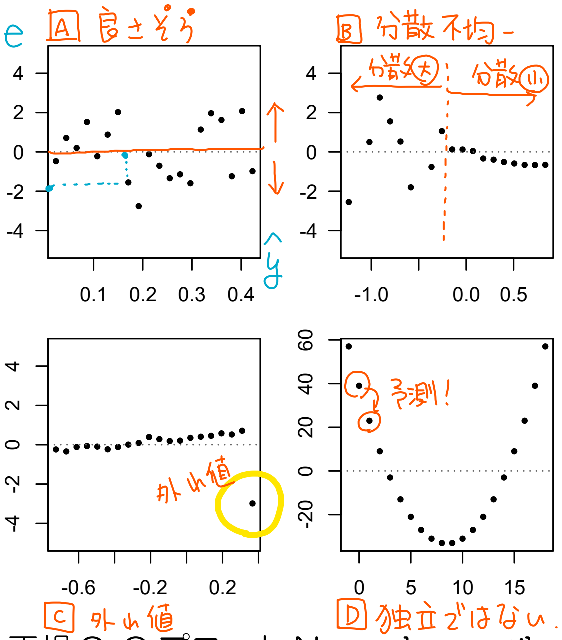
ノート１の（２）の仮定たちが間違っている可能性はないの？例えば、母集団に直線＋正規分布の関係を想定してきたけど、その仮定って正しかったの？



P 回帰分析の仮定
(1) 標本抽出
① 独立
② 同一分布
③ 無作為
(2) 母集団
① 正規分布
② 分散が均一

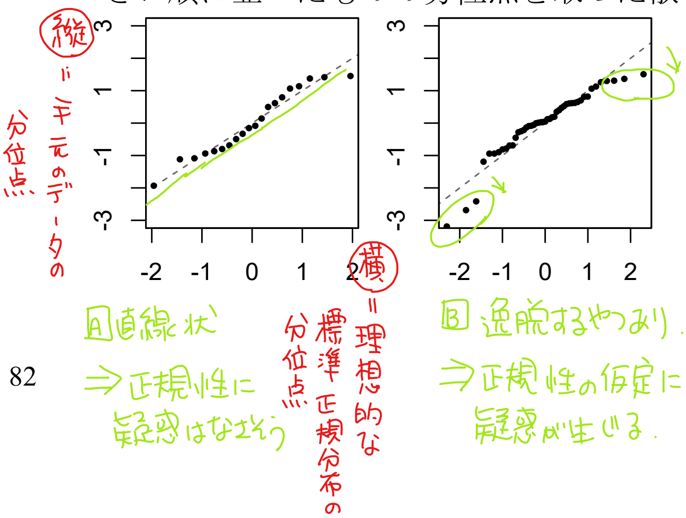
（１）残差プロット Residual plot

横軸に予測値$\hat{y}_i$、縦軸に残差$e_i = y_i - \hat{y}_i$を取った散布図。

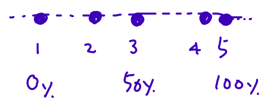


（２）正規 Q-Q プロット Normal quantile-quantile plot

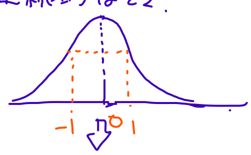
横軸に標準正規分布の分位点を、縦軸に残差を標準化し小さい順に並べたものの分位点を取った散布図。



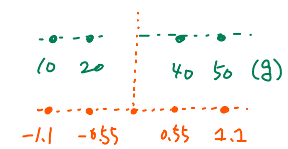
P 分位点
(参考) Lec2 順序統計量
① 離散的なとき



② 連続的なとき



P QQプロットの発想
① 標準化：標偏何個分？



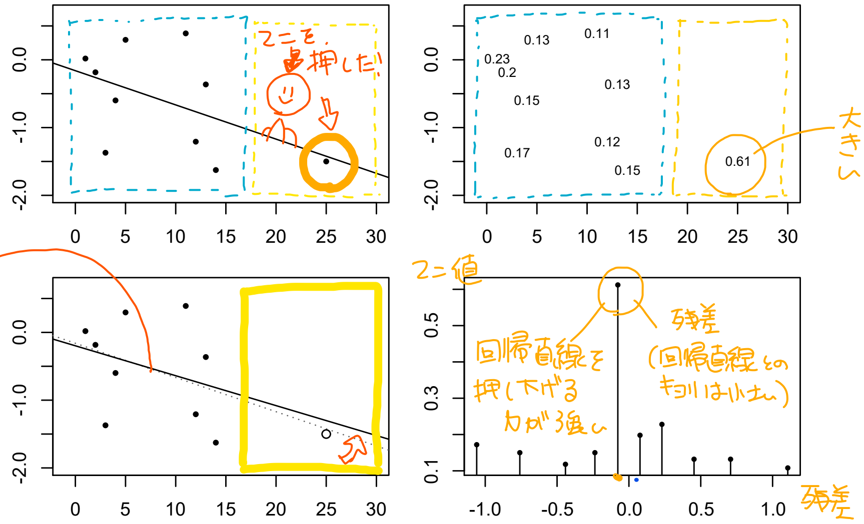
② 標準正規分布とデータを比較

Ⓟ てこ値の直観的意味

x軸方向において、外れてはっている点(=過疎地にいる点)は、「てこ」を「押しあげる/押しさげる」のように回帰直線に強い影響をもつ。

（３）てこ値 leverage

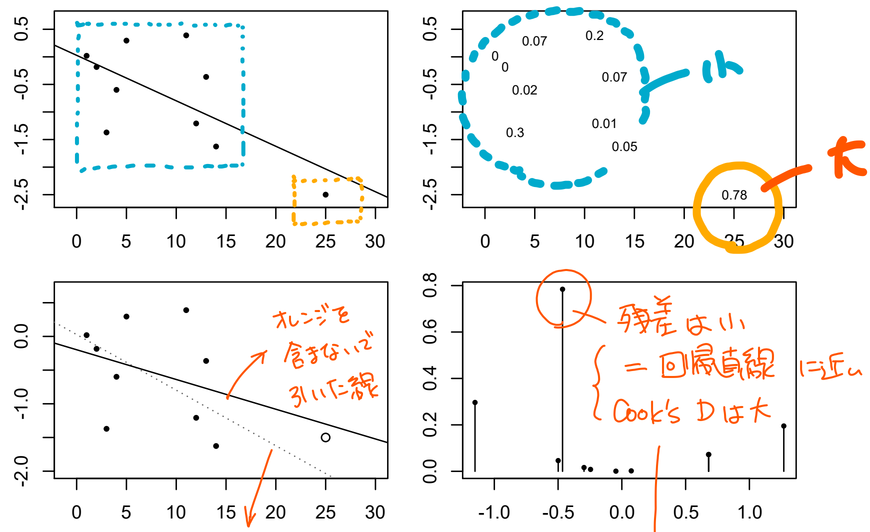
i 番目の観測値の回帰係数への影響度を表した指標。この値が大きいものは外れ値の候補となる。



$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i'=1}^{n}(x_{i'} - \bar{x})^2}$$

（４）クックの距離 Cook's distance

i 番目の観測値を除いて推定した場合に得られる回帰係数の値と、全観測値に基づく回帰係数のずれを表す指標。



そいつの有無して、結果が大きく変わる。

$$D_i = \frac{1}{p} r_i^2 \frac{h_i}{1 - h_i}$$

$D_i > 0.5$: i 番目の観測値は「影響力が大きめ」

$D_i > 1.0$: i 番目の観測値は「非常に大きな影響力」

資料

資料4-1　　最小二乗法による点推定

$$L(\hat{\beta}_0,\hat{\beta}_1)=\sum_{i=1}^{n}\left[y_i-(\hat{\beta}_0+\hat{\beta}_1 x_i)\right]^2$$

①$\hat{\beta}_0$に関して偏微分を行う

$$\begin{aligned}\frac{\partial}{\partial\hat{\beta}_0}L(\hat{\beta}_0,\hat{\beta}_1)&=\frac{\partial}{\partial\hat{\beta}_0}\sum_{i=1}^{n}\left[y_i-(\hat{\beta}_0+\hat{\beta}_1 x_i)\right]^2\\&=\sum_{i=1}^{n}\frac{\partial}{\partial\hat{\beta}_0}\left[y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right]^2\\&=\sum_{i=1}^{n}\frac{\partial}{\partial\hat{\beta}_0}\left[y_i^2-y_i\hat{\beta}_0-y_i\hat{\beta}_1 x_i-\hat{\beta}_0 y_i+\hat{\beta}_0^2+\hat{\beta}_0\hat{\beta}_1 x_i\right.\\&\qquad\left.-\hat{\beta}_1 x_i y_i+\hat{\beta}_1 x_i\hat{\beta}_0+\hat{\beta}_1^2 x_i^2\right]\\&=\sum_{i=1}^{n}\frac{\partial}{\partial\hat{\beta}_0}\left[\left(y_i^2-2\hat{\beta}_1 x_i y_i+\hat{\beta}_1^2 x_i^2\right)+2\hat{\beta}_0\left(-y_i+\hat{\beta}_1 x_i\right)+\hat{\beta}_0^2\right]\\&=\sum_{i=1}^{n}\left\{0+\frac{\partial}{\partial\hat{\beta}_0}\left[2\hat{\beta}_0\left(-y_i+\hat{\beta}_1 x_i\right)\right]+\frac{\partial}{\partial\hat{\beta}_0}\hat{\beta}_0^2\right\}\\&=\sum_{i=1}^{n}\left\{2\left(-y_i+\hat{\beta}_1 x_i\right)+2\hat{\beta}_0\right\}\\&=\sum_{i=1}^{n}\left\{2\left(-y_i+\hat{\beta}_1 x_i\right)\right\}+2n\hat{\beta}_0\end{aligned}$$

したがって、$\frac{\partial}{\partial\hat{\beta}_0}L(\hat{\beta}_0,\hat{\beta}_1)=0$となるのは、$\sum_{i=1}^{n}\left\{2\left(-y_i+\hat{\beta}_1 x_i\right)\right\}+2n\hat{\beta}_0=0$のとき。

$$\begin{aligned}\hat{\beta}_0&=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{\beta}_1 x_i\right)\\&=\frac{1}{n}\sum_{i=1}^{n}y_i-\hat{\beta}_1\frac{1}{n}\sum_{i=1}^{n}x_i\\&=\bar{y}-\hat{\beta}_1\bar{x}\end{aligned}$$

②$\hat{\beta}_1$に関して偏微分を行う

$$
\begin{aligned}
\frac{\partial}{\partial\hat{\beta}_1}L(\hat{\beta}_0,\hat{\beta}_1) &= \frac{\partial}{\partial\hat{\beta}_1}\sum_{i=1}^{n}[y_i-(\hat{\beta}_0+\hat{\beta}_1x_i)]^2 \\
&= \sum_{i=1}^{n}\frac{\partial}{\partial\hat{\beta}_1}[(y_i^2-2\hat{\beta}_0y_i+\hat{\beta}_0^2)-2\hat{\beta}_1(x_iy_i-\hat{\beta}_0x_i)+\hat{\beta}_1^2x_i^2] \\
&= \sum_{i=1}^{n}\left\{0-\frac{\partial}{\partial\hat{\beta}_1}[2\hat{\beta}_1(x_iy_i-\hat{\beta}_0x_i)]+\frac{\partial}{\partial\hat{\beta}_1}\hat{\beta}_1^2x_i^2\right\} \\
&= \sum_{i=1}^{n}\{-2x_i(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)\} \\
&= \sum_{i=1}^{n}\{-2x_i(y_i-\bar{y}-\hat{\beta}_1\bar{x}-\hat{\beta}_1x_i)\} \\
&= \sum_{i=1}^{n}\{-2x_i[(y_i-\bar{y})-\hat{\beta}_1(x_i-\bar{x})]\} \\
&= -2\sum_{i=1}^{n}\{x_i(y_i-\bar{y})\}+2\hat{\beta}_1\sum_{i=1}^{n}\{x_i(x_i-\bar{x})\}
\end{aligned}
$$

したがって、$\frac{\partial}{\partial\hat{\beta}_1}L(\hat{\beta}_0,\hat{\beta}_1)=0$となるのは、$-2\sum_{i=1}^{n}\{x_i(y_i-\bar{y})\}-2\hat{\beta}_1\sum_{i=1}^{n}\{x_i(x_i-\bar{x})\}$のとき。

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}\{x_i(y_i-\bar{y})\}}{\sum_{i=1}^{n}\{x_i(x_i-\bar{x})\}} \\
&= \frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2} \\
&= \frac{s_{xy}}{s_x^2}
\end{aligned}
$$