ひかり統計塾

統計検定２級
公式問題集(CBT対応版)の解説

カテゴリー10. 線形モデルの分野

10-1: 回帰分析の分野
(p142-159)

統計検定2級　CBT問題集　PART.2 目次

[C10-2] 分散分析の分野
↓
[C10-1] 回帰分析の分野
の順での学習をおすすめします
「回帰分析」で「分散分析」を
使いますので

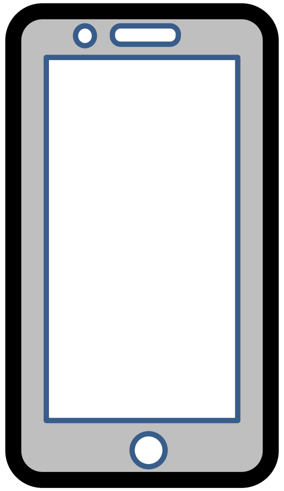
[C10]
[CATEGORY.10]
線形モデルの分野

[C10-1] 回帰分析の分野

[C10-2] 分散分析の分野

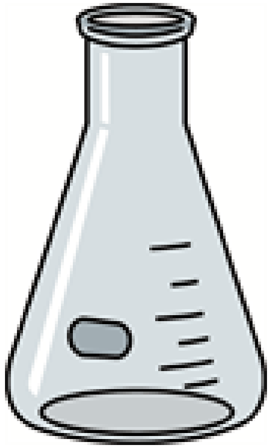
回帰分析の各問題の解説を
行う前に、「回帰分析の基礎」を
ご説明します。
回帰分析の基礎をご存知の方は、
(p142.1a)～(p142.1L)を
スキップし、
(p142.2)～を視聴ください

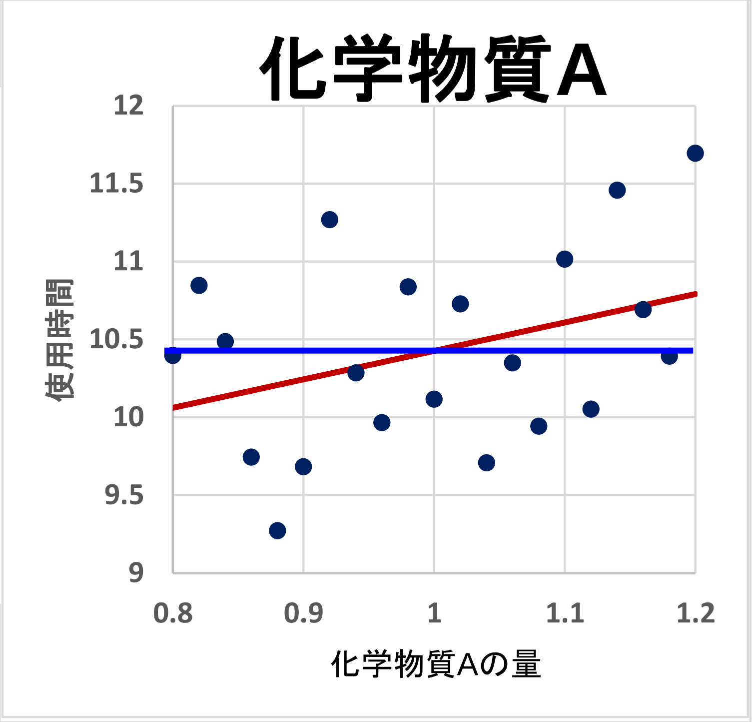
(p142.1a)[C10-1]回帰分析の基礎: 対象とする問題

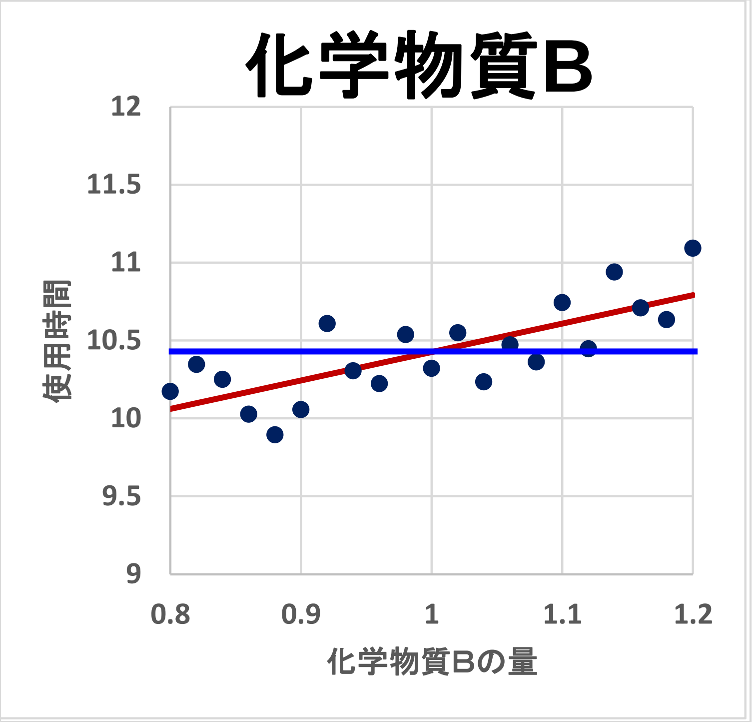


（問題）スマホの1充電当たりの使用時間を長くする検討を実施中と仮定。
・スマホのある素子作製工程で使う化学物質A、または化学物質Bの量を大きくすると、使用時間が長くなりそうな傾向がみられました。

(1)「化学物質の量(x)」と「使用時間(y)」の関係（式）は？　直線を想定すると？
(2)化学物質の量を変えると、使用時間が変わると言えるか？　　　を考えます



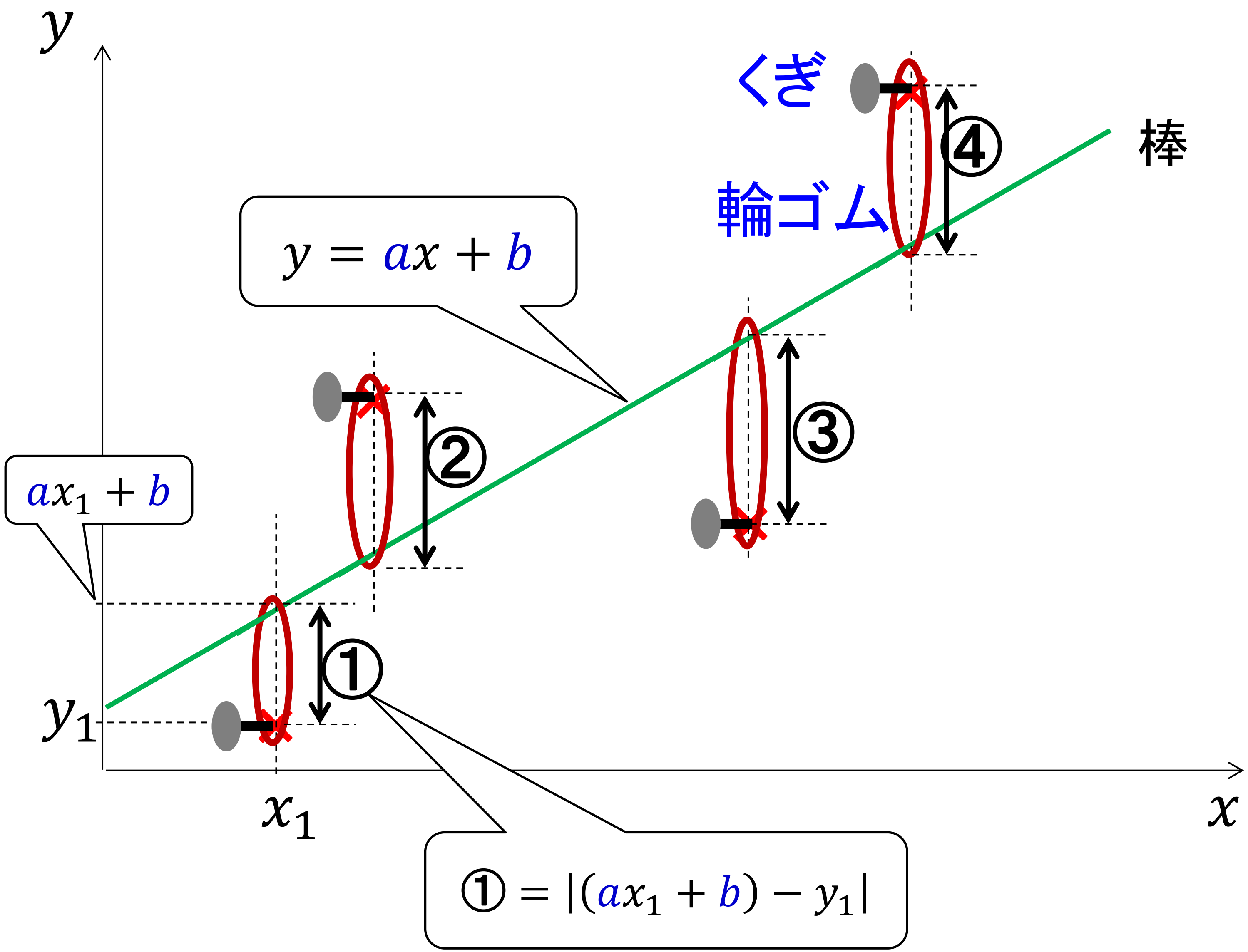




(1)直線を引く場合
直線をどう決める？
(回帰直線)

(2)傾き=0ですか？
傾き≠0ですか？

(p142.1b)[C10-1]回帰分析の基礎: 回帰直線の求め方



（大雑把な説明）
　左図で、釣り合う「棒」の
　位置を探す

（より正確な説明）
　$①^2 + ②^2 + ③^2 + ④^2$ を
最小にするように、
直線($y = ax + b$)を決める
⇒係数a, bを決める
「最小二乗法」と呼ばれます

(p142.1c)[C10-1]回帰分析の基礎: 回帰直線の式

以降、$y = ax + b$ → $y = \beta_1 x + \beta_0$ と書きます。

最小二乗法により決定

$y = \widehat{\beta_1} x + \widehat{\beta_0}$

$\widehat{\beta_1}, \widehat{\beta_0}$は、それぞれ$\beta_1, \beta_0$の「推定値」

公式(回帰直線、回帰係数):

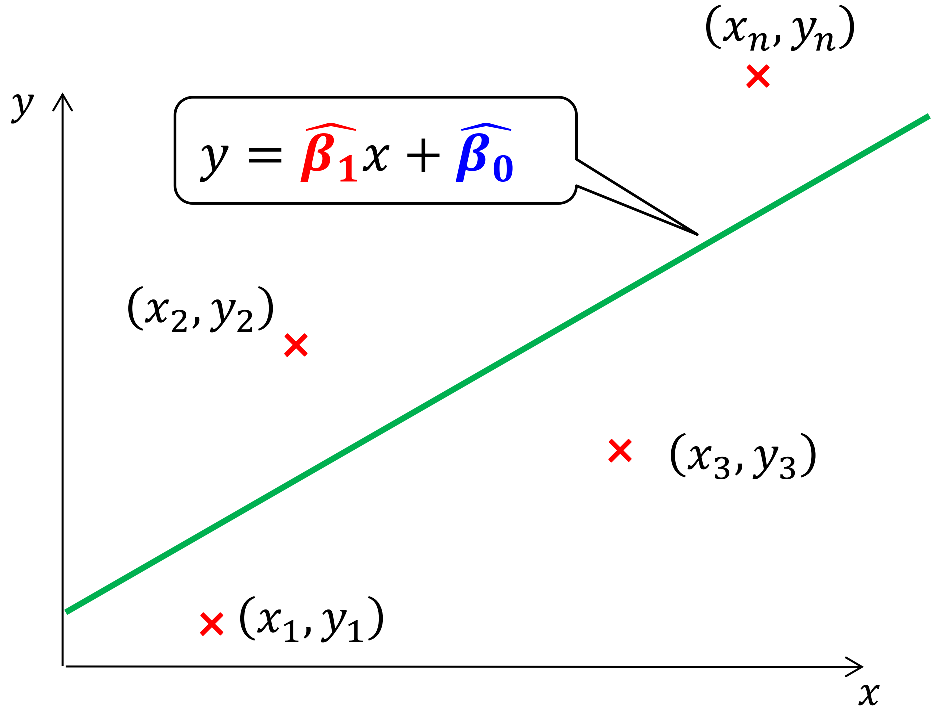
n個の点:$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ に対する回帰直線: $y = \widehat{\beta_1} x + \widehat{\beta_0}$

回帰係数: $\widehat{\beta_1} = \dfrac{S_{xy}}{S_{xx}}$, $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$

x, yの平均: $\bar{x} = \dfrac{1}{n}(x_1 + x_2 + \cdots + x_n)$, $\bar{y} = \dfrac{1}{n}(y_1 + y_2 + \cdots + y_n)$

偏差積和: $S_{xy} = (x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})$

偏差平方和: $S_{xx} = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2$



(p142.1d)[C10-1]回帰分析の基礎: 回帰直線の特徴,憶え方

$$y = \widehat{\beta_1}x + \widehat{\beta_0}, \quad \widehat{\beta_1} = \frac{S_{xy}}{S_{xx}}, \quad \widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$$

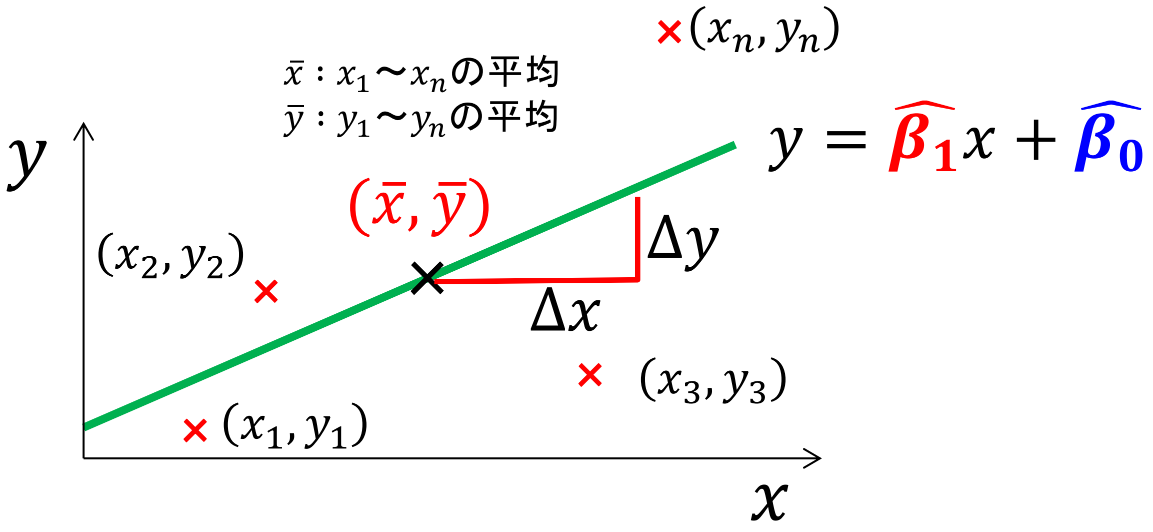
特徴①: 回帰直線は点$(\bar{x}, \bar{y})$を通る

役立つ問題例:
・回帰直線を選ぶ
(例:p143問1[2])

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$$

$$\bar{y} = \widehat{\beta_1}\bar{x} + \widehat{\beta_0}$$

点$(\bar{x}, \bar{y})$は回帰直線 $y = \widehat{\beta_1}x + \widehat{\beta_0}$ 上にある



特徴②: $\widehat{\beta_1}$は、「傾き」に対応

$$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}}$$

$$\widehat{\beta_1} \sim \frac{\Delta y}{\Delta x} \sim \frac{(y-\bar{y})}{(x-\bar{x})} \sim \frac{(x-\bar{x})(y-\bar{y})}{(x-\bar{x})(x-\bar{x})} \sim \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})(x-\bar{x})} \sim \frac{\sum_{i=1}^{n}(x_i-\bar{x})\boxed{(y_i-\bar{y})}}{\sum_{i=1}^{n}(x_i-\bar{x})\boxed{(x_i-\bar{x})}} = \frac{S_{xy}}{S_{xx}}$$

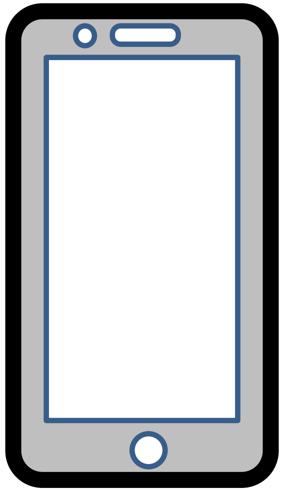
傾き

重み

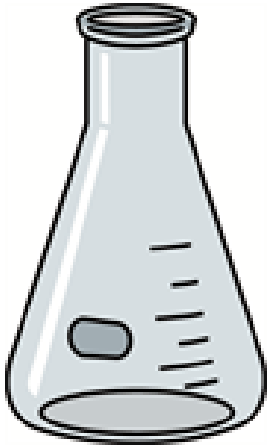
偏差積和: $S_{xy} = \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$

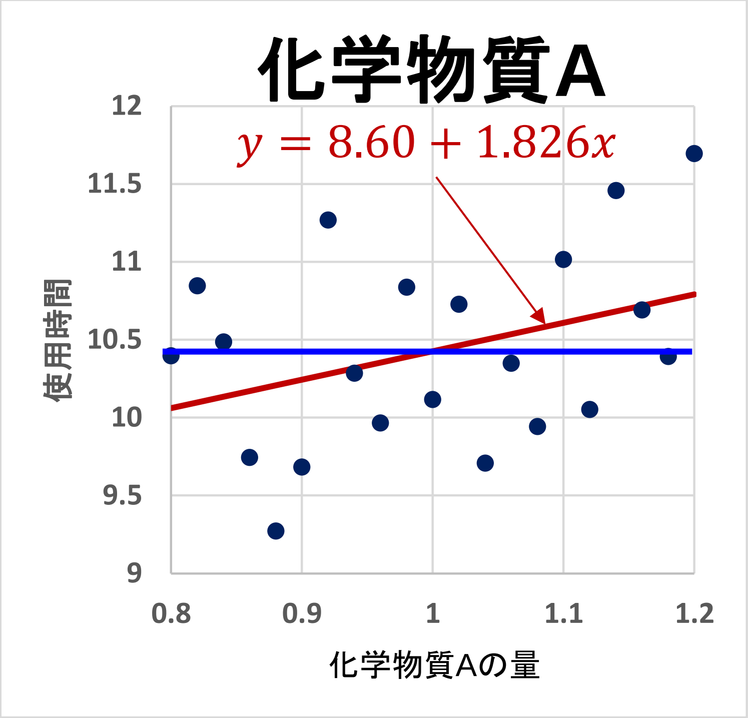
偏差平方和: $S_{xx} = \sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})$

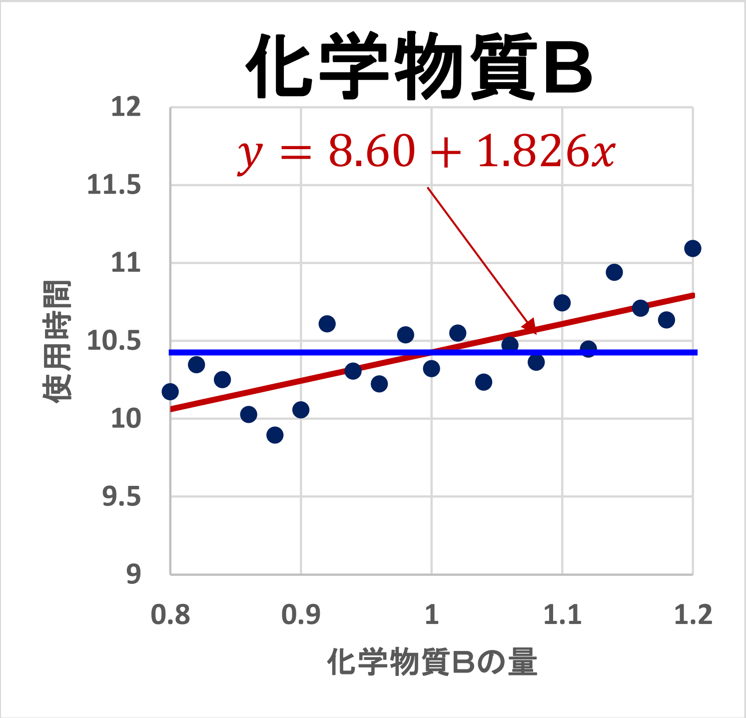
(p142.1e)[C10-1]回帰分析の基礎: 対象とする問題



(1)「化学物質の量(x)」と「使用時間(y)」の関係（式）は？　直線を想定すると？
(2)化学物質の量を変えると、使用時間が変わると言えるか？　　を考えます







済

(1)直線を引く場合
直線をどう決める？
⇒同じ回帰直線を得た

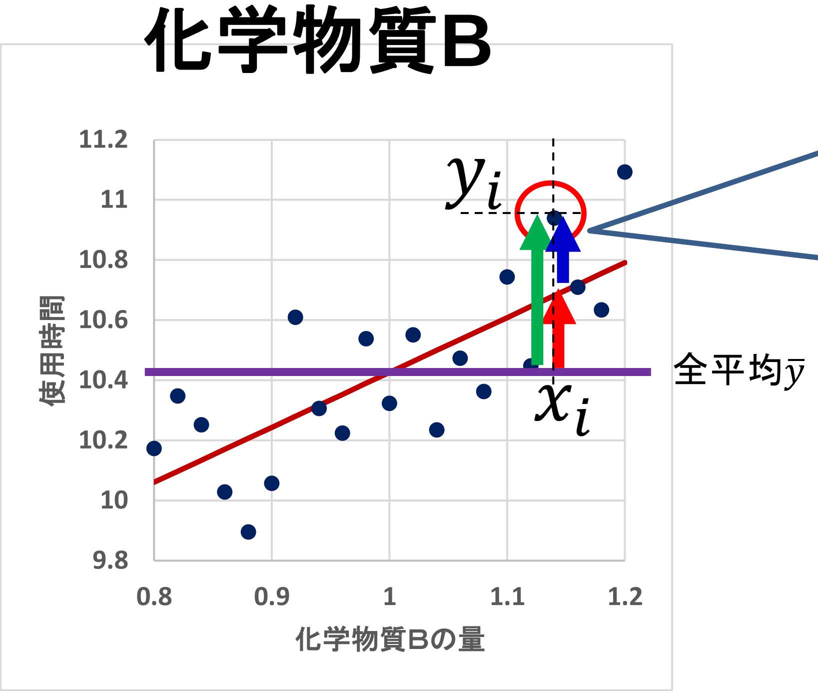
(2)傾き≠0と言える？

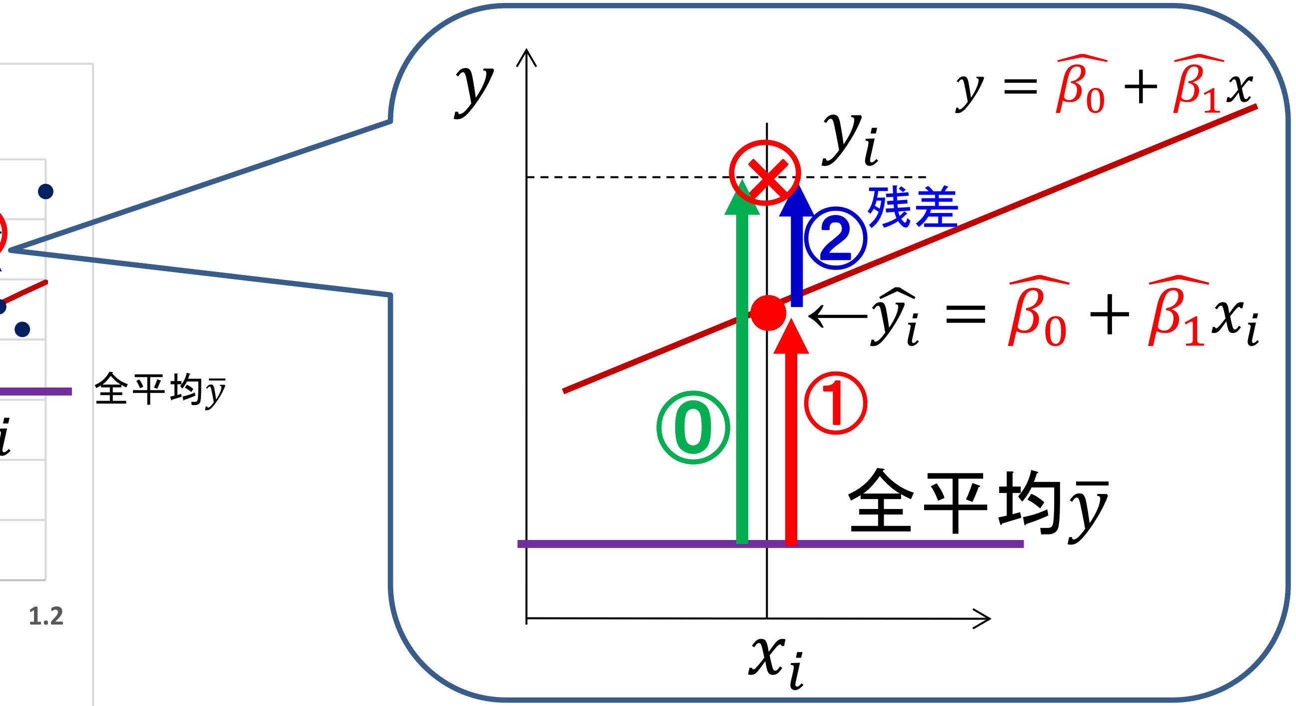
回帰の評価

$y = 8.60 + 1.826x$ ⇒ $x = 1.1$の時、$y = 10.61$と言い切れる？

(p142.1f)[C10-1]回帰分析の基礎: 回帰の評価、平方和の分解

分散分析の場合と同様にばらつきを①回帰、②残差の2つに分けます





(注)回帰では、
「誤差」の代わりに
「残差」を使います
(参考)入門統計解析法
p188 [注8.5]

⓪総平方和: $S_T = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\left((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})\right)^2$

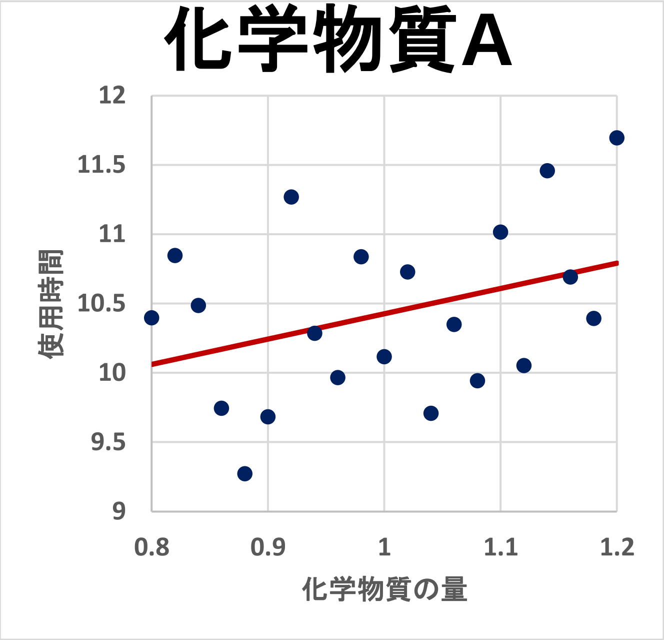
①回帰(Regression)による平方和: $S_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

②残差平方和: $S_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

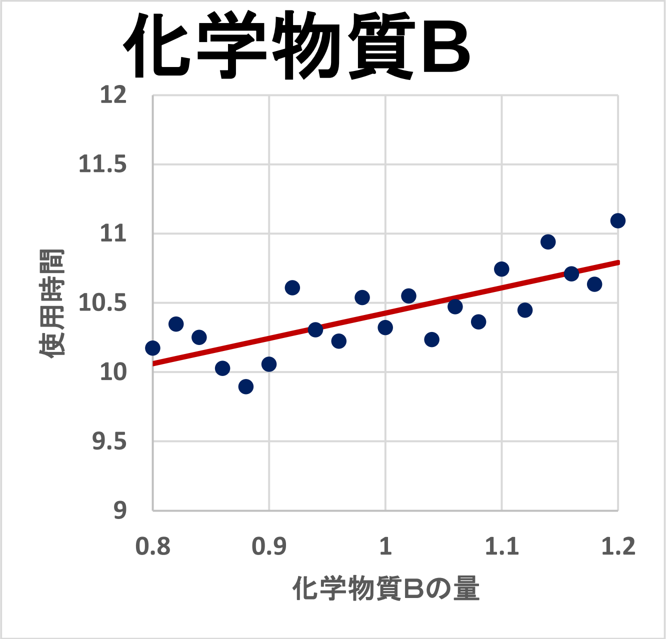
↓(参考)入門統計解析法
p186 (8.17)式

分散分析の場合と同様に $S_T = S_R + S_e$ が成立します！

(p142.1g)[C10-1]回帰分析の基礎: 決定係数(寄与率)



化学物質	A	B
回帰: S_R	1.027	1.027
残差: S_e	6.789	0.754
計: S_T	7.816	1.781
決定係数	0.131	0.577



①回帰 13.1%
②残差 86.9%
⓪総平方和

総平方和 S_Tのうち、
回帰の寄与 S_Rの割合が
決定係数(寄与率)です。

①回帰 57.7%
②残差 42.3%
⓪総平方和

化学物質	A	B
相関係数	0.362	0.759
相関係数^2	0.131	0.577
決定係数	0.131	0.577

$$決定係数 = \frac{S_R}{S_T} = r^2 \quad (r: 相関係数)$$

(p142.1h)[C10-1]回帰分析の基礎: 分散分析のステップ

理論:帰無仮説が正しい時、検定統計量: $F_0 = \frac{V_R}{V_e}$は自由度$(\boldsymbol{\phi_R}, \boldsymbol{\phi_e})$のF分布に従う

(1): 2つの仮説:
帰無仮説 H_0: $\beta_1 = 0$
対立仮説 H_1: $\beta_1 \neq 0$

回帰式:
$y = \beta_0 + \beta_1 x$

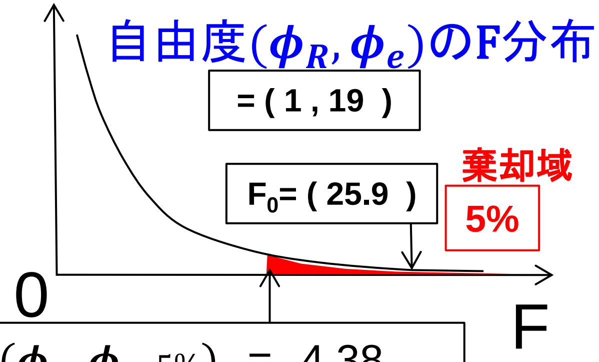
$\bar{x} = ($ 1.0000 $)$ $S_{xx} = ($ 0.3080 $)$
$\bar{y} = ($ 10.4261 $)$ $S_{xy} = ($ 0.5624 $)$

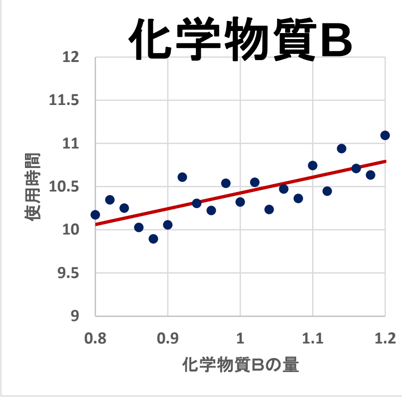
$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}}$
$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$

$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}} = 1.826$
$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$=8.600

(2): 有意水準αを定めます
α=0.05=5%

(3):自由度を求め、棄却域を定めます
データ総数n=(21),
全自由度$\boldsymbol{\phi_T} = n - 1 = ($ 20 $)$
回帰の自由度: (単回帰では) ⇒ $\boldsymbol{\phi_R} =$ 1
残差の自由度:$\boldsymbol{\phi_e} = \boldsymbol{\phi_T} - \boldsymbol{\phi_R}$ =(19)
棄却域:$F > F(\boldsymbol{\phi_R}, \boldsymbol{\phi_e}, 5\%)$
$F($ 1 ,(19), 5%$)$ =(4.38)





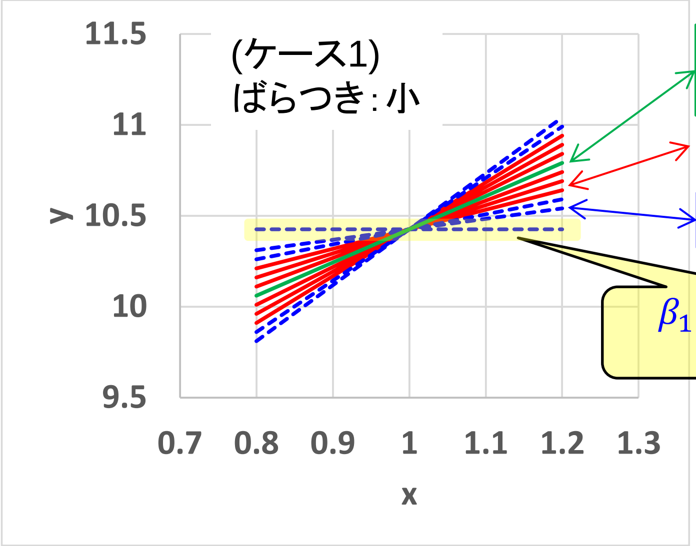
(4)分散分析表

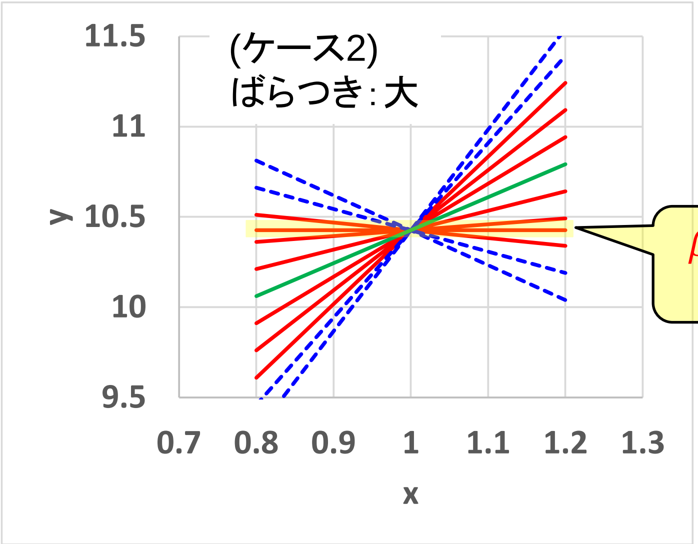
要因	(a)平方和 S		(b)自由度 φ		(c)平均平方 V=S/φ	F_0 =V_R/V_e	棄却域の下限
R(回帰)	1.027	÷	1	=	1.027	25.9	4.38
e(残差)	0.754	÷	19	=	0.0397	>	
計	1.781		20				

大小関係(<,>)

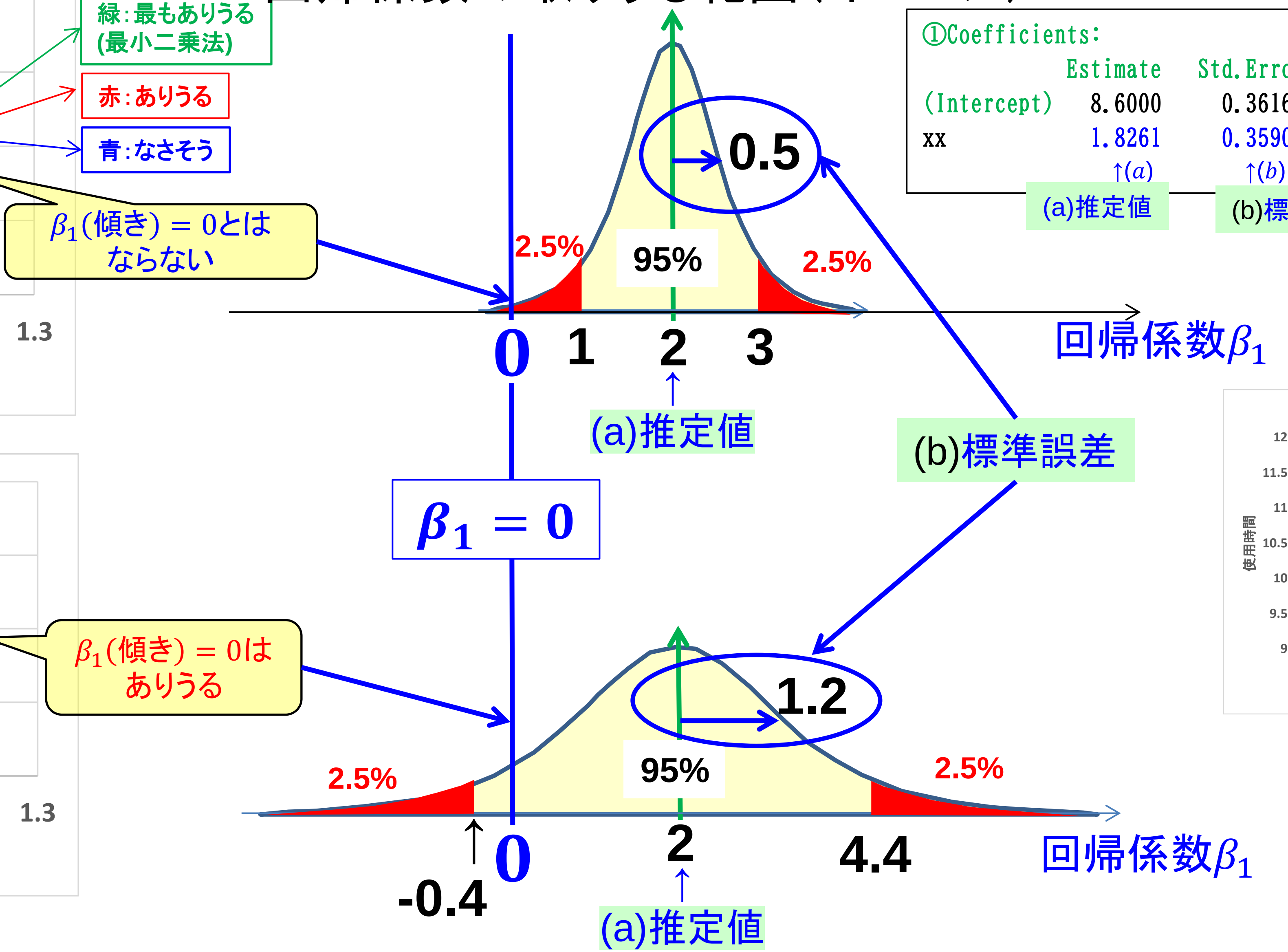
(5): H_0を棄却する・しないの判定
F_0 =(25.9) は棄却域に(ある ・ ない)
どっちですか?
(●) $F = F_0$が棄却域にあるので、H_0を棄却する
⇒ H_1は正しいと言える
( ) $F = F_0$が棄却域にないので、H_0は棄却できない
⇒ H_1は正しいかどうか何も言えない

(p142.1i)[C10-1]回帰分析の基礎：回帰係数≠0の検定



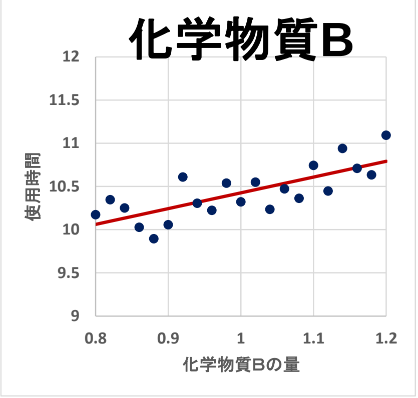


回帰係数の取りうる範囲（イメージ）




```
①Coefficients:
             Estimate   Std.Error t value Pr(>|t|)
(Intercept)    8.6000     0.3616   23.780 1.34e-15 ***
xx             1.8261     0.3590    5.086 6.56e-05 ***
               ↑(a)       ↑(b)
```

(a)推定値　(b)標準誤差

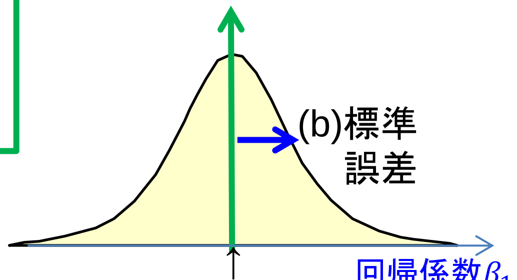


(p142.1j)[C10-1]回帰分析の基礎: 回帰係数≠0の検定

(公式) $t=\dfrac{(a)\text{推定値}-\beta_1}{(b)\text{標準誤差}}$ ~自由度ϕ_eのt分布

単回帰分析での残差の自由度$\phi_e=n-2$

(n: データ組数) $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$



注目する回帰係数: (例) β_1

(a)推定値(Estimate)=1.8261

(b)標準誤差(Std. Error)=0.3590

(c)残差の自由度$=21-2=19$

$$t=\frac{(a)\text{推定値}-\beta_1}{(b)\text{標準誤差}}=\frac{1.8261-\beta_1}{0.3590}\sim\text{自由度}\phi_e=19\text{の}t\text{分布}$$

統計ソフトRによる出力結果

```
①Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.6000     0.3616  23.780 1.34e-15 ***
xx             1.8261     0.3590   5.086 6.56e-05 ***
                ↑(a)        ↑(b)    ↑(e)
```

日本語版(ワンコピペエクセルシートより)

	回帰係数	①Estimate 推定値	①Std. Error 標準誤差	①t value 検定統計量(t 値)	①Pr(>|t|) P値	区間推定(下限)	区間推定(上限)
(①Intercept)(切片)	β_0 (^)=	8.6000	0.3616	23.7805	1.33227E-15	7.8430	9.3569
(傾き)	β_1 (^)=	1.8261	0.3590	5.0864	0.000066	1.0747	2.5775
		↑(a)	↑(b)	↑(e)		↑(g2)	↑(g1)

$y=\widehat{\beta_1}x+\widehat{\beta_0}=1.8261\,x+8.6$

検定

(1) 帰無仮説(H_0): $\beta_1=0$
対立仮説(H_1): $\beta_1\neq 0$

両側検定

(2) 有意水準 α=0.05

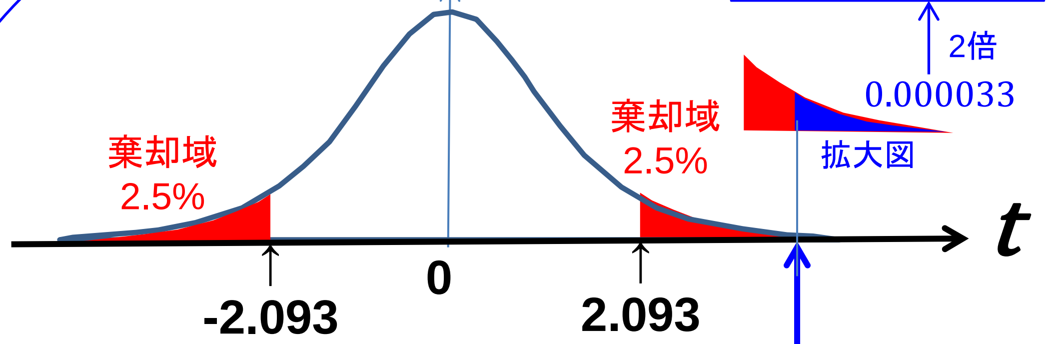
(3) 棄却域 |t|≧2.093

(4) 帰無仮説($\beta_1=0$)が正しい時の検定統計量

$$t=\frac{1.8261-\beta_1}{0.3590}=\frac{1.8261-0}{0.3590}=5.0864$$

(5) 検定統計量(t値)は棄却域にある
⇒帰無仮説(H_0)を棄却する⇒対立仮説(H_1)($\beta_1\neq 0$)は正しいと言える

自由度$\phi_e=19$のt分布



t値 = 5.0864

(p142.1k)[C10-1]回帰分析の基礎: 回帰係数の区間推定

前ページと同じ

(公式) $t = \frac{(a)\text{推定値} - \beta_1}{(b)\text{標準誤差}} \sim$ 自由度ϕ_eのt分布

単回帰分析での残差の自由度$\phi_e = n - 2$

(n: データ組数) $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

注目する回帰係数: β_1

(a)推定値(Estimate)=1.8261

(b)標準誤差(Std. Error)=0.3590

(c)残差の自由度$= 21 - 2 = 19$

$$t = \frac{(a)\text{推定値} - \beta_1}{(b)\text{標準誤差}} = \frac{1.8261 - \beta_1}{0.3590} \sim \text{自由度}\phi_e = 19\text{の}t\text{分布}$$

統計ソフトRによる計算結果

```
①Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.6000     0.3616  23.780 1.34e-15 ***
xx             1.8261     0.3590   5.086 6.56e-05 ***
                ↑(a)       ↑(b)     ↑(e)
```

日本語版(ワンコピペエクセルシートより)

①Coefficients	回帰係数	①Estimate 推定値	①Std. Error 標準誤差	①t value 検定統計量 (t 値)	①Pr(>|t|) P値	区間推定 (下限)	区間推定 (上限)
(①Intercept) (切片)	β_0 (^)=	8.6000	0.3616	23.7805	1.33227E-15	7.8430	9.3569
(傾き)	β_1 (^)=	1.8261	0.3590	5.0864	0.000066	1.0747	2.5775
		↑(a)	↑(b)	↑(e)		↑(g2)	↑(g1)

区間推定の場合

$t \sim$ 自由度$\phi_e = 19$のt分布 の時、信頼確率95%で

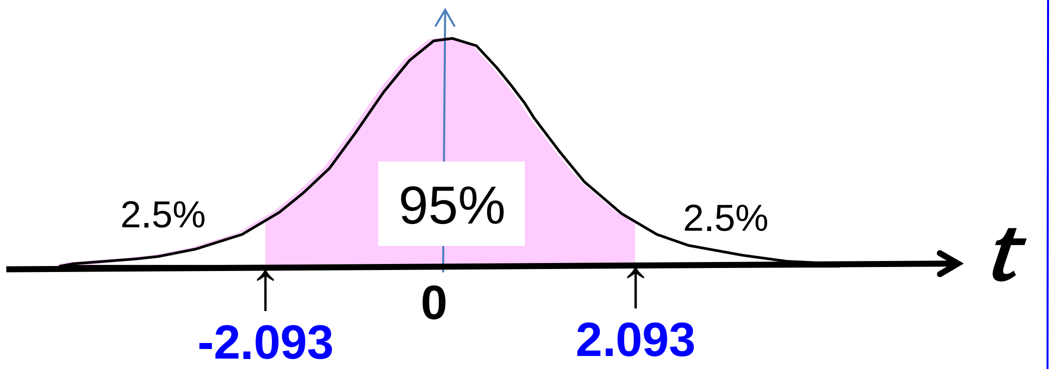
$$|t| \leq 2.093$$

$$|t| = \left|\frac{1.8261 - \beta_1}{0.3590}\right| \leq 2.093$$

$$|1.8261 - \beta_1| \leq 2.093 \times 0.3590 = 0.7514$$

$$1.0747 \leq \beta_1 \leq 2.5775$$

自由度$\phi_e = 19$のt分布



(p142.1L)[C10-1]回帰分析の基礎: 統計ソフト「R」の結果

統計検定2級の回帰の問題では、統計ソフトウェア(R)による出力結果が示され、
これに基づく解釈に関する問題が頻繁に出されます　⇒読み取れるようにしておいてください

(問題例)CBT問題集p147,148,151,171,188

統計ソフトウェア:Rによる回帰分析結果

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.31152 -0.14160 -0.01048  0.13554  0.32984
```

残差の
最小値･最大値、
四分位数･中央値
(関連)CBT問題集
p152問3[1]④

```
①Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.6000     0.3616  23.780 1.34e-15 ***
xx            1.8261     0.3590   5.086 6.56e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

②Residual standard error: 0.1992 on 19 degrees of freedom
③Multiple R-squared:  0.5766, Adjusted R-squared:  0.5543
④F-statistic: 25.87 on 1 and 19 DF,  p-value: 6.558e-05
```

日本語版:
(ワンコピペエクセルシートでの回帰分析結果)

分散分析表			↓検定統計量			↓(④p-value)	↓(1:効果あり、0:無)	
要因	平方和	自由度	平均平方	F_0値	棄却域(下限)	P値	判定	効果
R(回帰)	1.027	1	1.027051962	25.8710	4.3807	6.55798E-05	1	有
E(残差)	0.754	19	0.039698956	ーーーー	ーーーー	ーーーー	ーーーー	ーーーー
計	1.781	20	ーーーー	ーーーー	ーーーー	ーーーー	ーーーー	ーーーー

残差の標準誤差($\sqrt{V_e}$)=　0.199245969　↑④F-statistic　④回帰の自由度(DF)=　1

④残差の自由度(DF)=　19

●ステップ4ーーーーーー:　検定・推定を行う

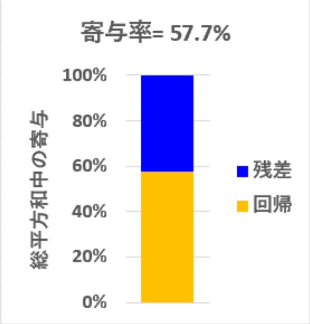
検定：　検定統計量F_0値は棄却域に　あるので

帰無仮説　H_0：「回帰に意味はない」は　棄却できる

対立仮説　H_1：「回帰に意味がある」　と言える

P値=　0.00007　⇔ α：0.05

(別解)P値はα(有意水準)より　小さいのでH0を棄却できる



●ステップ5ーーーーーー:　回帰分析の結果：

決定係数(寄与率)=　0.576564　(③Multiple R-squared)

自由度調整済決定係数=　0.5542779　(③Adjusted R-squared)

残差の標準誤差($\sqrt{Ve}$)=　0.199246　(②Residual standard error)　②残差の自由度(degrees of freedom)=　19

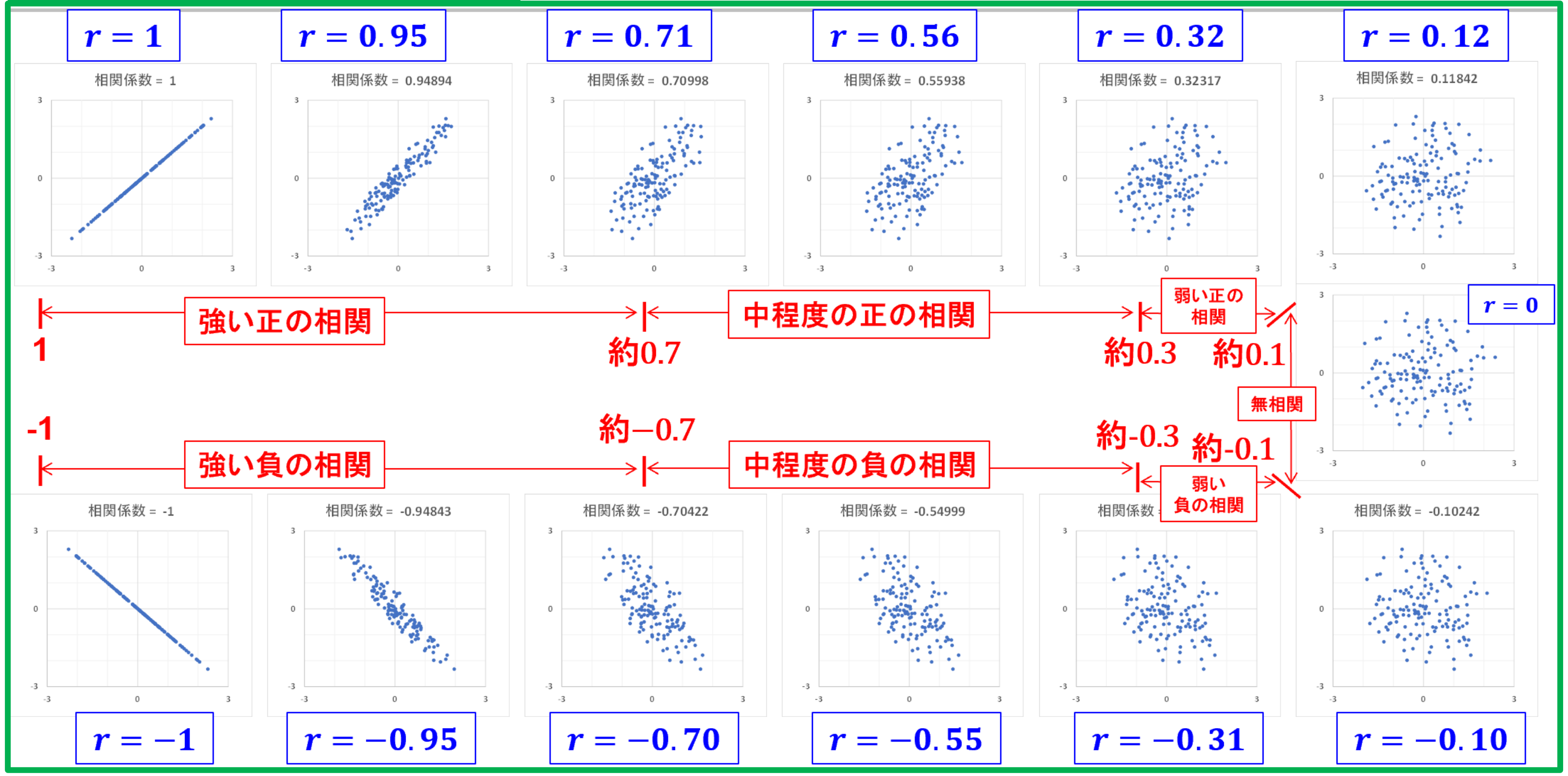
		①Estimate	①Std. Error	①t value	①Pr(>|t|)		
	①Coefficients			検定統計量		区間推定	区間推定
	回帰係数	点推定値	標準誤差	(t 値)	P値	(下限)	(上限)
(①Intercept)(切片)	β_0(^)=	8.6000	0.361639519	23.7805	1.33227E-15	7.8430	9.3569
(傾き)	β_1(^)=	1.8261	0.359016316	5.0864	6.55798E-05	1.0747	2.5775
	$y=\beta_0(\hat{})+\beta_1(\hat{})x$				(比較値：0)		

(p142.2) [C10-1]問1[1]. 最小二乗法・傾きの検定

[1]:ABランク
[2]:ABランク
[3]:Bランク

小問[1]:　散布図⇒相関係数を選ぶ問題

散布図と相関係数の例



[1]
←左の例を参考にすると、

・p142の散布図：負の相関がある
⇒①②(正の相関)は×

③ほど($r=0.094$)弱くはない

⑤ほどの強い負の相関($r=-0.994$)はない

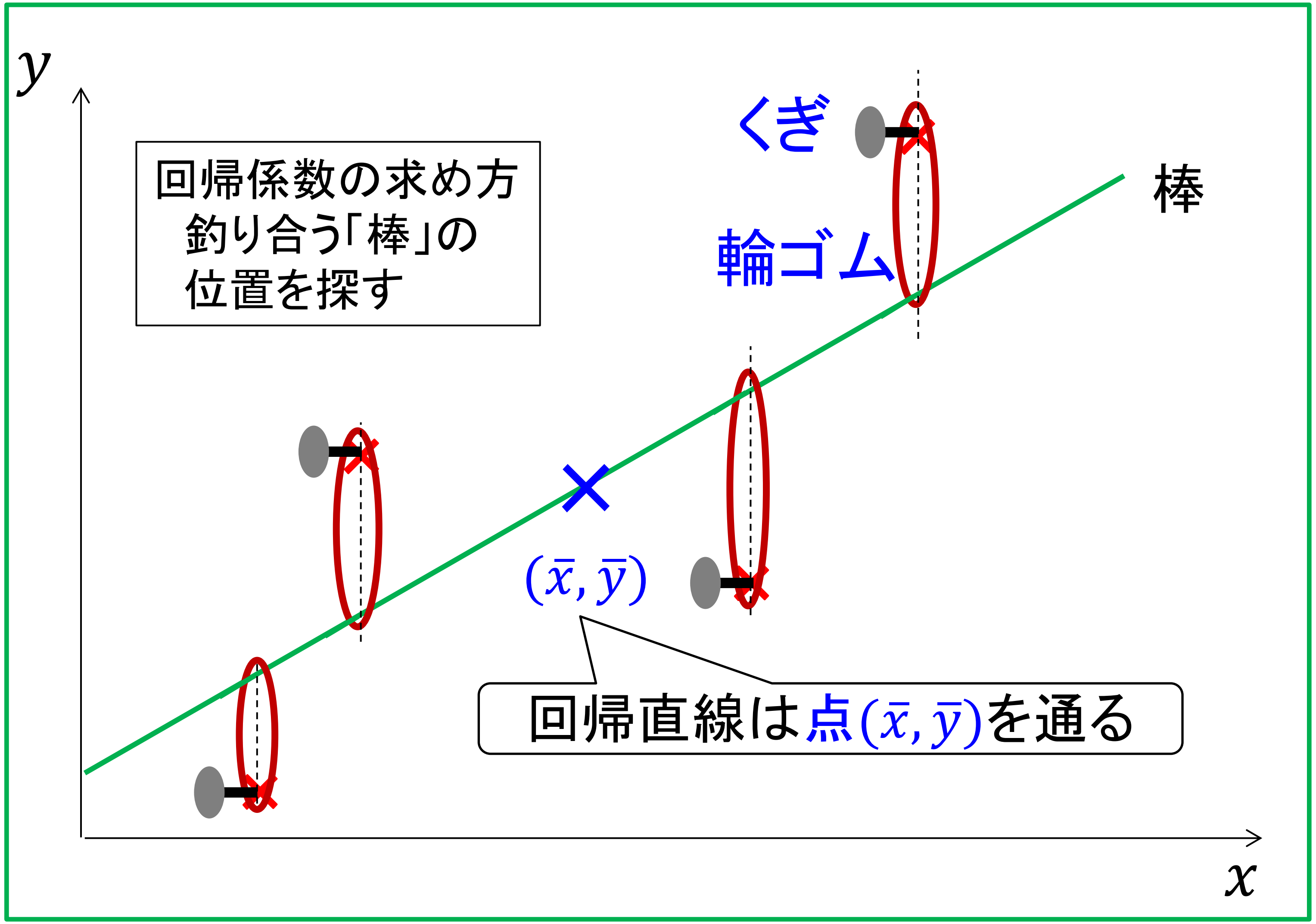
⇒④が適当

[1] (答)④

(p142.3) [C10-1]問1[2]. 最小二乗法・傾きの検定

[1]:ABランク
[2]:ABランク
[3]:Bランク

小問[2]:　回帰直線を選ぶ問題



①:下に寄り過ぎ。
(左上の点から離れすぎ)
③: 左下側に寄り過ぎ。
④: 散布図の右下付近の点からずれ過ぎ。
⑤: 直線の右上側には3点しかなく、
バランスが悪い。

⇒②がバランス面でも、
$(\bar{x}, \bar{y})$を通りそうという観点からも
最もいい。

[2](答)②

(p142.4) [C10-1]問1[3]. 最小二乗法・傾きの検定

[1]:ABランク
[2]:ABランク
[3]:Bランク

小問[3]:　回帰式：$y = \beta_1 x + \beta_0$

回帰係数: β_1
(a)推定値(Estimate)= -0.1450
(b)標準誤差(Std. Error)= 0.02316

(c)残差の自由度$\phi_e =$　$n - 2 = 23$

但し、データ組数 $n =$　$2015 - 1991 + 1 = 25$

(1) 帰無仮説(H_0): $\beta_1 = 0$

帰無仮説(H_0)が正しい時の
検定統計量の値は、

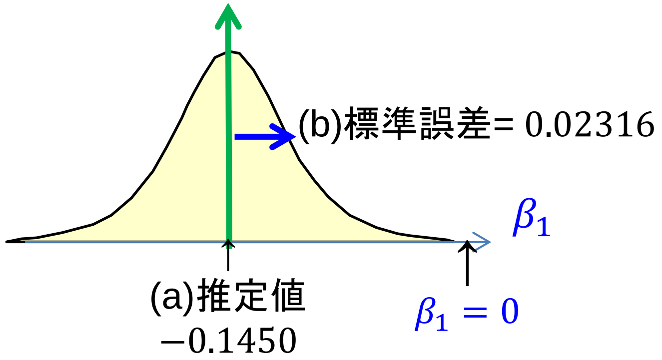
★ $t = \dfrac{(a)推定値 - \beta_1}{(b)標準誤差} = \dfrac{-0.1450 - \beta_1}{0.02316} = \dfrac{-0.1450 - 0}{0.02316} = -6.2651$

★ 検定統計量が従う分布は・・・ 自由度ϕ_e =23のt分布

(公式) $t = \dfrac{(a)推定値 - \beta_1}{(b)標準誤差}$ ～自由度ϕ_eのt分布

単回帰分析での残差の自由度: $\phi_e = n - 2$

(n: データ組数) $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$



仮に、有意水準5%で、
「対立仮説$\beta_1 \neq 0$」として検定すると、
棄却域|t|≧2.069 ⇒ t値は棄却域内にある
⇒H_0を棄却⇒H_1:正しい、 となります。

済

[3]検定統計量(t値)= -6.27
t～自由度23のt分布に従う
[3](答)⑤

(p142.4) [C10-1]問1[3]. 最小二乗法・傾きの検定

[1]:ABランク
[2]:ABランク
[3]:Bランク

小問[3]:　　回帰式：$y = \beta_1 x + \beta_0$

回帰係数: β_1
(a)推定値(Estimate)= -0.1450
(b)標準誤差(Std. Error)= 0.02316

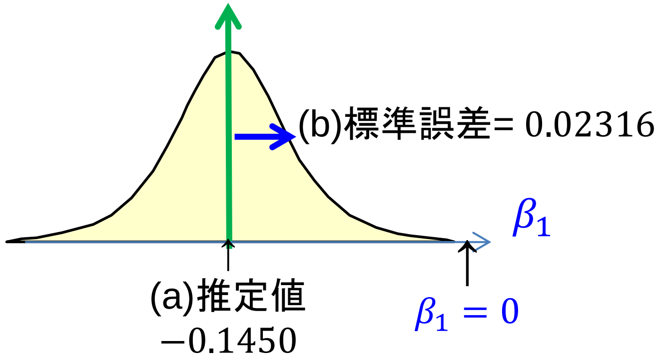
(c)残差の自由度$\phi_e =$　$n - 2 = 23$

但し、データ組数 $n =$　$2015 - 1991 + 1 = 25$

(公式) $t = \dfrac{(a)\text{推定値} - \beta_1}{(b)\text{標準誤差}}$ ～自由度ϕ_eのt分布

単回帰分析での残差の自由度: $\phi_e = n - 2$

(n: データ組数) $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$



(1) 帰無仮説(H_0): $\beta_1 = 0$

帰無仮説(H_0)が正しい時の
検定統計量の値は、

★ $t = \dfrac{(a)\text{推定値} - \beta_1}{(b)\text{標準誤差}} = \dfrac{-0.1450 - \beta_1}{0.02316} = \dfrac{-0.1450 - 0}{0.02316} = -6.2651$

★ 検定統計量が従う分布は･･･ 自由度ϕ_e =23のt分布

仮に、有意水準5%で、
「対立仮説$\beta_1 \neq 0$」として検定すると、
棄却域|t|≧2.069 ⇒ t値は棄却域内にある
⇒H_0を棄却⇒H_1:正しい、 となります。

済

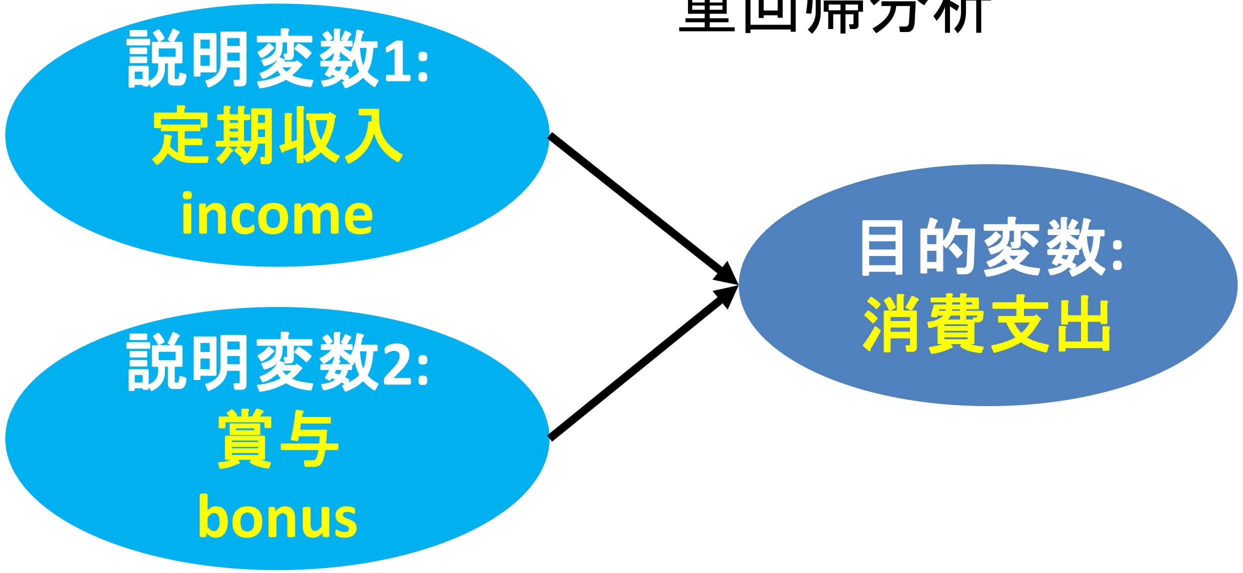
[3]検定統計量(t値)= -6.27
t～自由度23のt分布に従う
[3](答)⑤

(p146.1) [C10-1]問2[1]. 重回帰結果の解釈・単回帰予測

[1]:Bランク
[2]:Bランク

小問[1]

重回帰分析



回帰式:

消費支出 ＝ ● ＋ ■ × 定期収入 ＋▲× 賞与
（推定値）

の具体的な式はどうなりますか？

回帰式:

消費支出 ＝ 14.58851 ＋ 0.39461 × 定期収入 ＋ 0.47247 × 賞与

それぞれの単位は「万円」

(例) 回帰式の参照箇所

```
. . .
①Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.6000     0.3616  23.780 1.34e-15 ***
x             1.8261     0.3590   5.086 6.56e-05 ***
---
. . .
```

単回帰分析: $y = 8.6000 + 1.8261x$

```
. . .
①Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.6000     0.3616  23.780 1.34e-15 ***
x             1.8261     0.3590   5.086 6.56e-05 ***
y             2.8000     0.1590   (略)
---
. . .
```

重回帰分析: $y = 8.6000 + 1.8261x + 2.8000y$

(p146.2) [C10-1]問2[1]. 重回帰結果の解釈・単回帰予測

[1]:Bランク
[2]:Bランク

小問[1]:

回帰式

$$14.58851 + 0.39461 \times 定期収入 + 0.47247 \times 賞与 = 消費支出$$

消費支出, 定期収入, 賞与の
それぞれの単位は「万円」

①$14.58851 + 0.39461 \times (定期収入+1) + 0.47247 \times 賞与 = 消費支出 + \boxed{0.39461}$ ○
正しい

②$14.58851 + 0.39461 \times (定期収入+1) + 0.47247 \times (賞与 + 1) = 消費支出 + \boxed{0.39461 + 0.47247}$ ×
約0.39 ではない

③$14.58851 + 0.39461 \times 定期収入 \times (1 + 0.01) + 0.47247 \times 賞与 = 消費支出 + \boxed{0.39461 \times 定期収入 \times 0.01}$ ×

$消費支出 \times \frac{0.39}{100}$ ではない

④$14.58851 + 0.39461 \times 定期収入 \times (1 + 0.01) + 0.47247 \times 賞与 \times (1 + 0.01) = 消費支出 + \boxed{0.39461 \times 定期収入 \times 0.01 + 0.47247 \times 賞与 \times 0.01}$

$消費支出 \times \frac{0.39}{100}$ ではない ×

⑤定期収入:1万円増えたら消費支出:約0.39万円増える :○・・・①
定期収入:1%増えたら消費支出:約0.39%増える :×・・・③ なので、×

[1](答)①

(p146.3) [C10-1]問2[2]. 重回帰結果の解釈・単回帰予測

[1]:Bランク
[2]:Bランク

小問[2]:

I　$\bar{\hat{y}} = \bar{y}$ ですか？　(但し、$\bar{\hat{y}}$は、$x = x_i$での予測値：$\hat{y}_i$の平均)

●$x = x_i$での予測値：$\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i$

その平均は、

$$\bar{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = \frac{1}{n}\sum_{i=1}^{n} (\widehat{\beta_0} + \widehat{\beta_1} x_i) = \widehat{\beta_0} + \widehat{\beta_1}\bar{x} = \bar{y}$$

iには関係ない定数

I：正しい

II　世帯主収入合計(x)の平均: $\bar{x} = 41.0$　ですか？

●回帰式：$y = \widehat{\beta_1} x + \widehat{\beta_0}$

回帰係数の推定値：$\widehat{\beta_1} = 0.4121$, $\widehat{\beta_0} = 14.3931$

$\Rightarrow y = 0.4121x + 14.3931$

$\bar{y} = \widehat{\beta_1}\bar{x} + \widehat{\beta_0}$　$\Rightarrow$　$\bar{y} = 0.4121\bar{x} + 14.3931$

$\bar{\hat{y}} = \bar{y} = 31.3$ なので、

$31.3 = 0.4121\bar{x} + 14.3931$　$\Rightarrow$　$\bar{x} = 41.026$

II：正しい

公式(回帰直線、回帰係数)：

●回帰式：$y = \widehat{\beta_1} x + \widehat{\beta_0}$

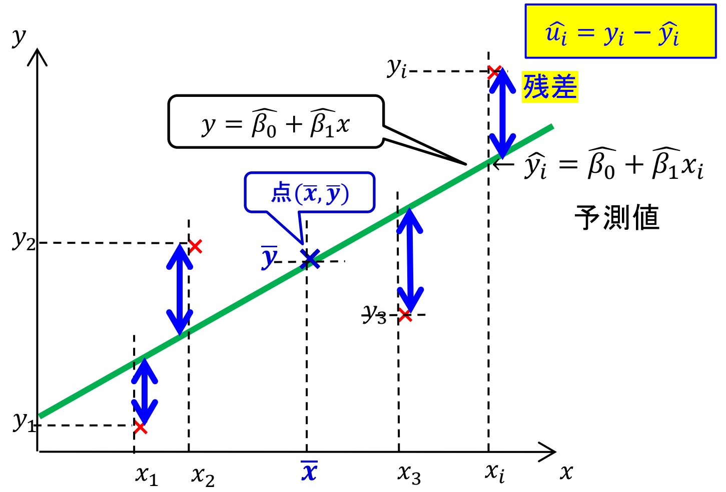
●回帰係数：$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}}$,　$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$

回帰直線は
点$(\bar{x}, \bar{y})$を通る

●x, yの平均：$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$,　$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

●偏差積和：$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

●偏差平方和：$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$



●$x = x_i$での予測値：$\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i$

●残差：$\hat{u}_i = y_i - \hat{y}_i = y_i - (\widehat{\beta_0} + \widehat{\beta_1} x_i)$

●残差の総和：$\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$

「残差の総和は0」←できれば憶えてください

(p146.4) [C10-1]問2[2]. 重回帰結果の解釈・単回帰予測

[1]:Bランク
[2]:Bランク

小問[2]:

III　予測値$\widehat{y_i}$ ＋残差$\widehat{u_i}$ ＝元のデータy_i ですか？

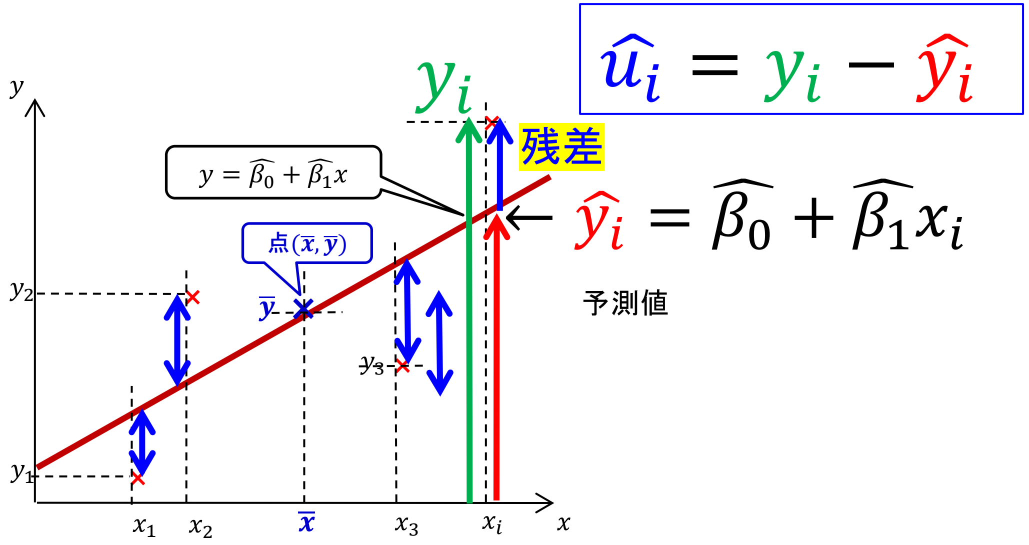
●$x = x_i$での予測値：　$\widehat{y_i}$

●残差：　$\widehat{u_i} = y_i - \widehat{y_i}$

足すと

$$\widehat{y_i} + \widehat{u_i} = \widehat{y_i} + (y_i - \widehat{y_i}) = y_i$$

III：正しい



前のページ分も含めて

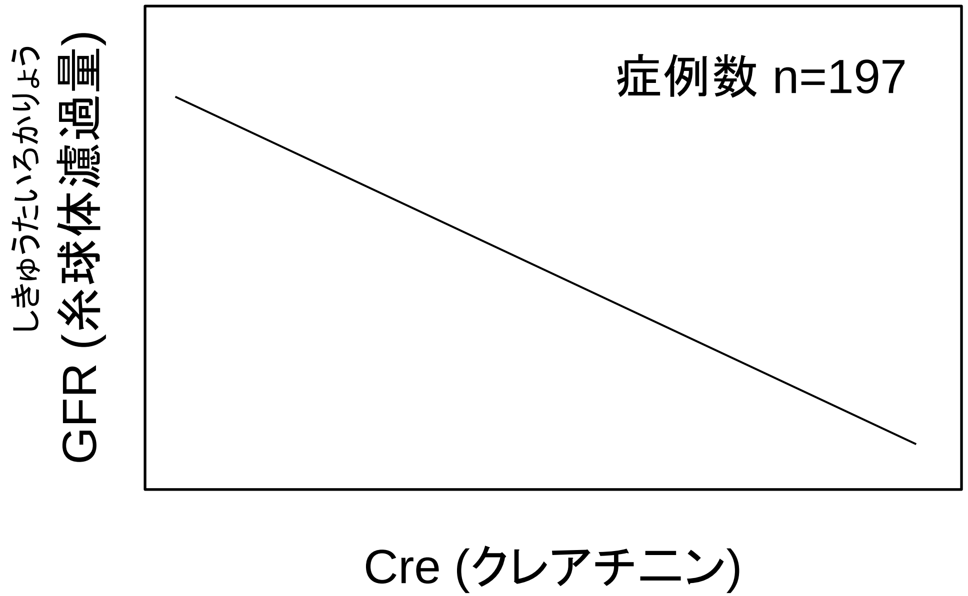
I　：正しい
II　：正しい
III：正しい

[2](答)⑤

済

(p151.1) [C10-1]問3. 出力結果の解釈・残差・信頼区間

[1]:BCランク
[2]:ABランク
[3]:BCランク



(p151.2) [C10-1]問3[1]. 出力結果の解釈・残差・信頼区間

[1]:BCランク
[2]:ABランク
[3]:BCランク

小問[1]:

①:正しいですか？

データ総数:$n = 197$
全自由度$\boldsymbol{\phi}_T = n - 1 = 196$
回帰の自由度: (単回帰なので) $\boldsymbol{\phi}_R = 1$
残差の自由度:$\boldsymbol{\phi}_e = \boldsymbol{\phi}_T - \boldsymbol{\phi}_R = 196 - 1 = 195$
p151下段の出力結果、下から3行目に「195 degrees of freedom」あり

(解説によると)
回帰モデルの自由度＝残差の自由度=195　であり、
決定係数とは関係ない⇒　　①は正しくない

②:正しいですか？

「決定係数」「自由度調整(修正)済決定係数」と
「判断が正しいかどうかの確率」とは直接関係ない
⇒　　②は正しくない

決定係数(寄与率) $= \dfrac{S_R}{S_T} = 1 - \dfrac{S_e}{S_T} = r^2$

S_R :回帰による平方和$= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
S_e :残差平方和$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
S_T :総平方和$= S_R + S_e = \sum_{i=1}^{n}(y_i - \bar{y})^2$
r: 相関係数(重回帰の場合、重相関係数)

自由度調整(修正)済決定係数 $= \mathbf{1} - \dfrac{S_e/\phi_e}{S_T/\phi_T}$

重回帰分析で
「意味なく説明変数増やすと決定係数は1に近づく」
という問題回避のためにつかわれるもの
「決定係数」より小さい値となる

回帰分析の自由度:
データ総数:nの時、
全自由度$\boldsymbol{\phi}_T = n - 1$
回帰の自由度: (単回帰では) $\boldsymbol{\phi}_R = 1$
（説明変数がp個の重回帰では、$\boldsymbol{\phi}_R = p$)

残差の自由度:$\boldsymbol{\phi}_e = \boldsymbol{\phi}_T - \boldsymbol{\phi}_R$

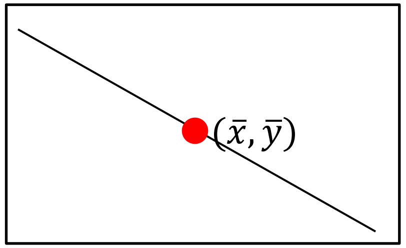
(p151.3) [C10-1]問3[1]. 出力結果の解釈・残差・信頼区間

[1]:BCランク
[2]:ABランク
[3]:BCランク

小問[1]の③

目的変数

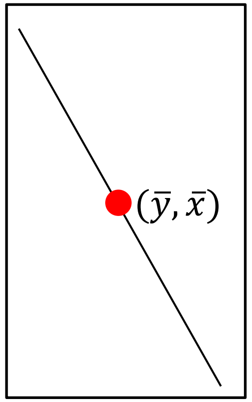
y (GFR)



x (Cre) 説明変数

目的変数

x (Cre)



y (GFR) 説明変数

公式

●回帰係数: $\widehat{\beta_1} = \dfrac{S_{xy}}{S_{xx}}$

x:説明変数
y:目的変数

●偏差積和: $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

●偏差平方和:

$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2, S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$

●相関係数: $r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$

相関係数が$r = \pm 1$の時のみ、

$\widehat{\beta_1} \times \widehat{\beta_1}' = 1, \quad \dfrac{1}{\widehat{\beta_1}} = \widehat{\beta_1}'$ となる

$$y = \widehat{\beta_1}x + \widehat{\beta_0}$$

$$x = \widehat{\beta_1}'y + \widehat{\beta_0}'$$

$$\boxed{\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}}}$$

$$\boxed{\widehat{\beta_1}' = \frac{S_{yx}}{S_{yy}} = \frac{S_{xy}}{S_{yy}}}$$

$$\widehat{\beta_1} \times \widehat{\beta_1}' = \frac{S_{xy}}{S_{xx}} \times \frac{S_{xy}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx}S_{yy}} = r^2$$

$$x = \frac{1}{\widehat{\beta_1}}y - \frac{1}{\widehat{\beta_1}}\widehat{\beta_0}$$

$$\frac{1}{\widehat{\beta_1}} = \widehat{\beta_1}' = -\frac{1}{60.263}$$ ですか？ ⇒この問題では、No!

$r^2 = 0.4888$（$p151$ 出力結果の下から2行目）なので、

$\widehat{\beta_1}' \neq \dfrac{1}{\widehat{\beta_1}} (= -0.0166)$です。 $\widehat{\beta_1}' = \dfrac{r^2}{\widehat{\beta_1}} = -\dfrac{0.4888}{60.263} = -0.0081$

$\boxed{\widehat{\beta_1} = -60.263}$

③は正しくない

$p154$の[1]
③の解説の
最後と一致

(p151.4) [C10-1]問3[1]. 出力結果の解釈・残差・信頼区間

[1]:BCランク
[2]:ABランク
[3]:BCランク

小問[1]の④

問題: 残差の平均 > 残差の中央値(-0.816)
⇒正しいですか？

```
Residuals:
    Min      1Q  Median      3Q     Max
-33.650 -10.471  -0.816   8.079  44.298
```

残差の四分位数 (箱ひげ図を思い出してください)

Min	1Q	Median (中央値)	3Q	Max
-33.650	-10.471	-0.816	8.079	44.298

(注)図は正確ではありません

●残差の総和: $\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$

「残差の総和は0」なので、

残差の平均 $= \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = 0$ > 残差の中央値

残差の中央値 $= -0.816$

⇒ 残差の平均(0) > 残差の中央値(-0.816)

●$x = x_i$での予測値: $\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i$

●残差: $\hat{u}_i = y_i - \hat{y}_i = y_i - (\widehat{\beta_0} + \widehat{\beta_1} x_i)$

●残差の総和: $\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$

「残差の総和は0」←できれば憶えてください

④は正しい

(p151.5) [C10-1]問3[1]. 出力結果の解釈・残差・信頼区間

[1]:BCランク
[2]:ABランク
[3]:BCランク

小問[1]の⑤

問3の例ではありません

統計ソフトウェア：Rによる回帰分析結果

```
①Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.6000     0.3616  23.780 1.34e-15 ***
xx             1.8261     0.3590   5.086 6.56e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

②Residual standard error: 0.1992 on 19 degrees of freedom
③Multiple R-squared:  0.5766, Adjusted R-squared:  0.5543
④F-statistic: 25.87 on 1 and 19 DF,  p-value: 6.558e-05
```

F-statisticは、「分散分析」における検定統計量です。

回帰係数が、有意水準1%で有意かどうかは、
Pr(>|t|)により判断できます。

⑤は正しくない

日本語版：
(ワンコピペエクセルシートでの回帰分析結果)

分散分析表　　↓検定統計量　　↓(④p-value)　　↓(1:効果あり、0:無)

要因	平方和	自由度	平均平方	F_0値	棄却域(下限)	P値	判定	効果
R(回帰)	1.027	1	1.027051962	25.8710	4.3807	6.55798E-05	1	有
E(残差)	0.754	19	0.039698956	――――	――――	――――	――――	――――
計	1.781	20	――――	――――	――――	――――	――――	――――

残差の標準誤差($\sqrt{V_e}$)= 0.199245969　↑④F-statistic　④回帰の自由度(DF)= 1

④残差の自由度(DF)= 19

●ステップ4――――――: 検定・推定を行う

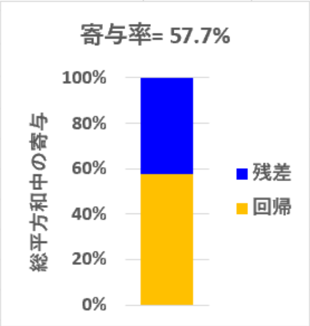
検定：　検定統計量F_0値は棄却域に あるので

帰無仮説 H_0：「回帰に意味はない」は 棄却できる

対立仮説 H_1：「回帰に意味がある」 と言える

P値= 0.00007 ⇔ α: 0.05

(別解)P値はα(有意水準)より 小さいのでH0を棄却できる



●ステップ5――――――: 回帰分析の結果：

決定係数(寄与率)= 0.576564 (③Multiple R-squared)

自由度調整済決定係数= 0.5542779 (③Adjusted R-squared)

残差の標準誤差($\sqrt{Ve}$)= 0.199246 (②Residual standard error)　②残差の自由度(degrees of freedom)= 19

①Coefficients 回帰係数		①Estimate 点推定値	①Std. Error 標準誤差	①t value 検定統計量 (t 値)	①Pr(>|t|) P値	区間推定 (下限)	区間推定 (上限)
(①Intercept)(切片)	β_0(^)=	8.6000	0.361639519	23.7805	1.33227E-15	7.8430	9.3569
(傾き)	β_1(^)=	1.8261	0.359016316	5.0864	6.55798E-05	1.0747	2.5775
	$y=\beta_0(\hat{})+\beta_1(\hat{})x$				(比較値：0)		

まとめ:①②③⑤は正しくない、④は正しい

[1](答)④

(p151.6) [C10-1]問3[2]. 出力結果の解釈・残差・信頼区間

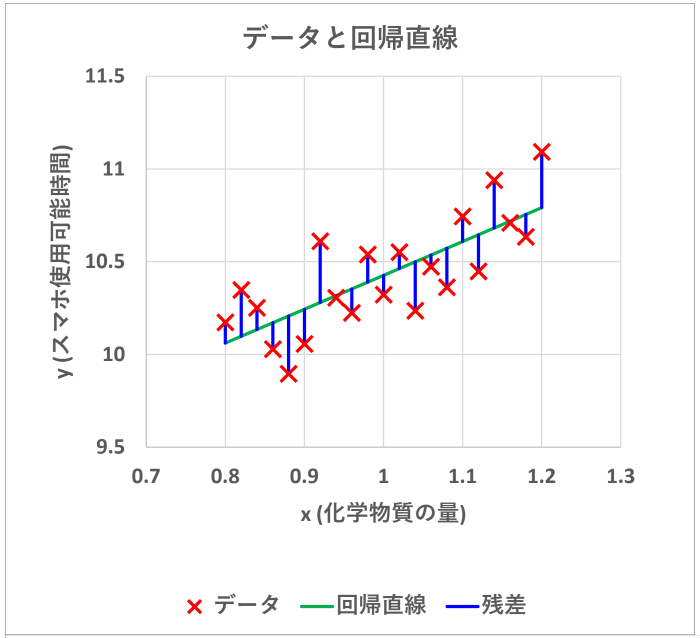
[1]:BCランク
[2]:ABランク
[2]:BCランク

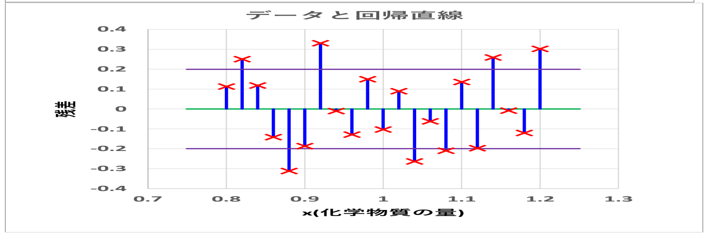
小問[2] :残差プロット・・・①～④から選ぶ問題です。

右の例の様に、残差プロットは、
元の散布図を反映しているので、
端部の値や、回帰直線から離れた点などに着目し、
よく見比べて判断しましょう

p151 右下の端の5個の点に着目する
・Cre=1.4より大きい3点 :　回帰直線より上にある
・Cre=1.4の2点 :　　　　回帰直線の少し上と、縦軸10程度分下にある
⇒これに合致するのは、①のみ

残差プロットの例
(問3とは違うケースです)





[2](答)①

(p151.7) [C10-1]問3[3]. 出力結果の解釈・残差・信頼区間

[1]:BCランク
[2]:ABランク
[3]:BCランク

小問[3] : Creの回帰係数の区間推定

(公式) $t = \dfrac{(a)推定値 - \beta_1}{(b)標準誤差}$ ～自由度ϕ_eのt分布

単回帰分析での残差の自由度$\phi_e = n - 2$

(n: データ組数) $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

注目する回帰係数: β_1

(a)推定値(Estimate)= -60.263

(b)標準誤差(Std. Error)= 4.414

(c)残差の自由度= 195 （症例数 n=197を使用)

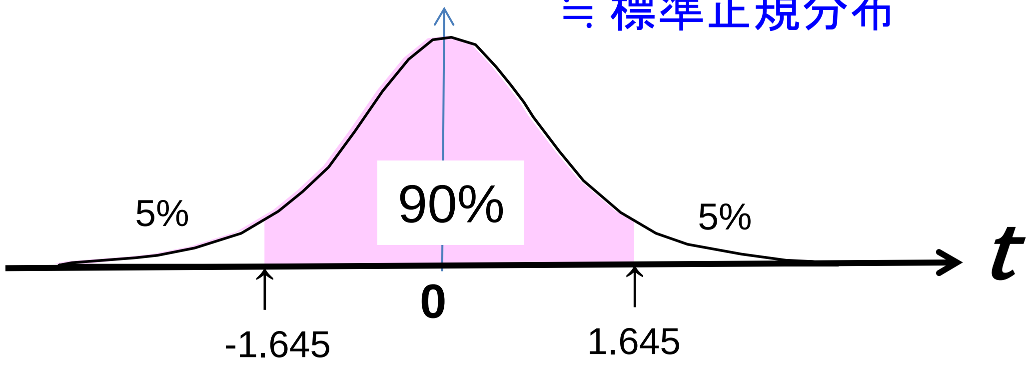
$$t = \frac{(a)推定値 - \beta_1}{(b)標準誤差} = \frac{-60.263 - \beta_1}{4.414}$$ ～自由度ϕ_e = (195)のt分布

t～自由度ϕ_e = (195)のt分布 の時、信頼確率90%で

$|t| \le$ (1.645)

自由度ϕ_e = (195)のt分布

≒ 標準正規分布



$$|t| = \left|\frac{-60.263 - \beta_1}{4.414}\right| \le 1.645$$

$$|-60.263 - \beta_1| \le 1.645 \times 4.414 = 7.26103$$

$$-67.524 \le \beta_1 \le -53.002$$

「回帰分析の基礎でご紹介した例」

t～自由度ϕ_e = 19のt分布 の時、信頼確率95%で

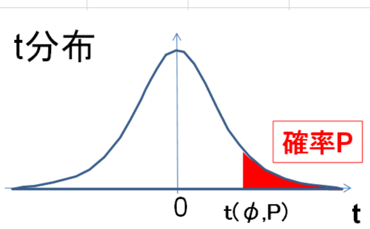
$|t| \le 2.093$

$$|t| = \left|\frac{1.8261 - \beta_1}{0.3590}\right| \le 2.093$$

$$|1.8261 - \beta_1| \le 2.093 \times 0.3590 = 0.7514$$

$$1.0747 \le \beta_1 \le 2.5775$$

	片側P=	0.100	0.05	0.025	0.01	0.005
	(両側2P=)	0.200	0.10	0.05	0.02	0.01
φ=	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	60	1.296	1.671	2.000	2.390	2.660
	120	1.289	1.658	1.980	2.358	2.617
	240	1.285	1.651	1.970	2.342	2.596
	∞	1.282	$\boxed{1.645}$	1.960	2.326	2.576



[3](答)②

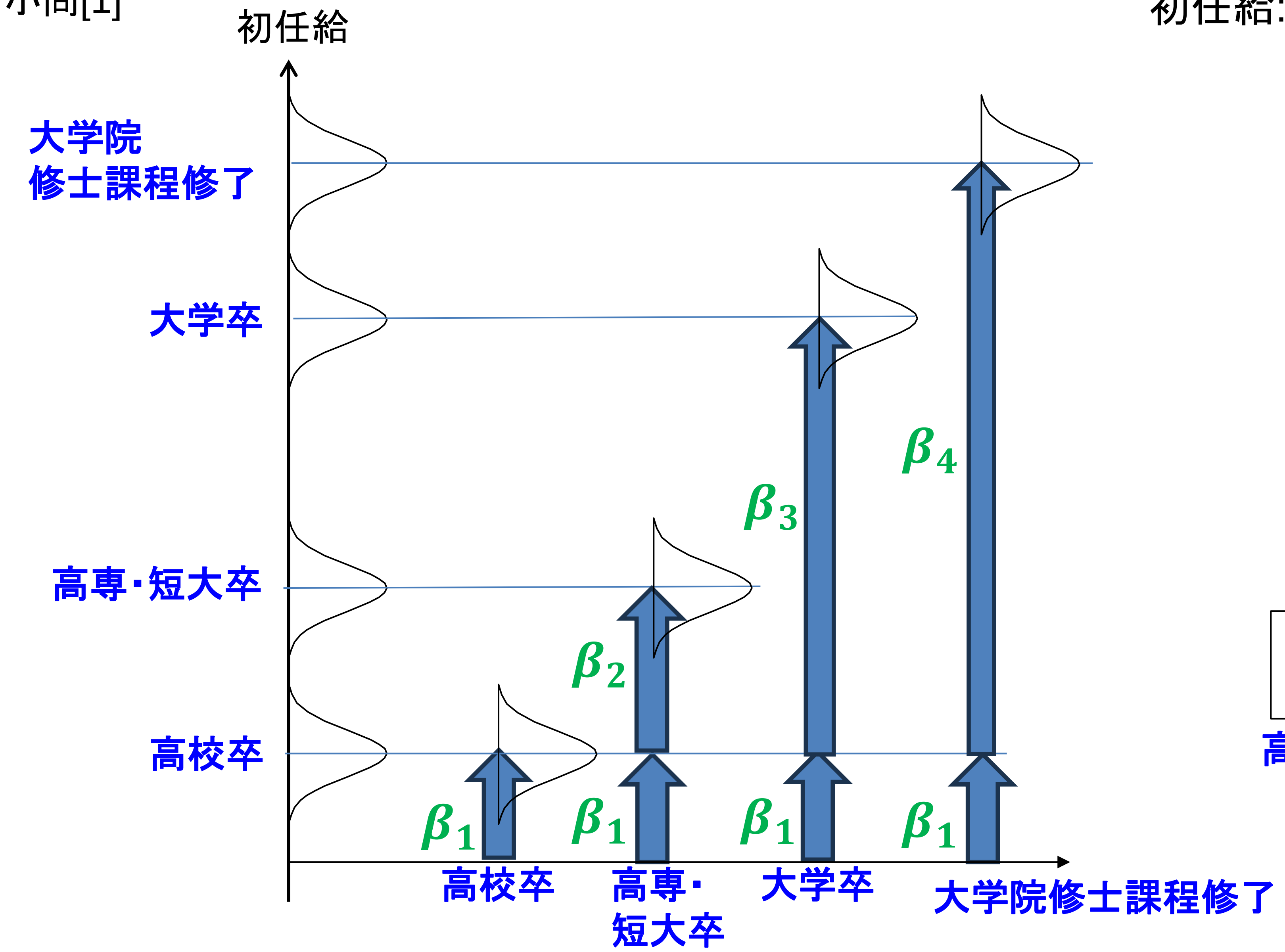
済

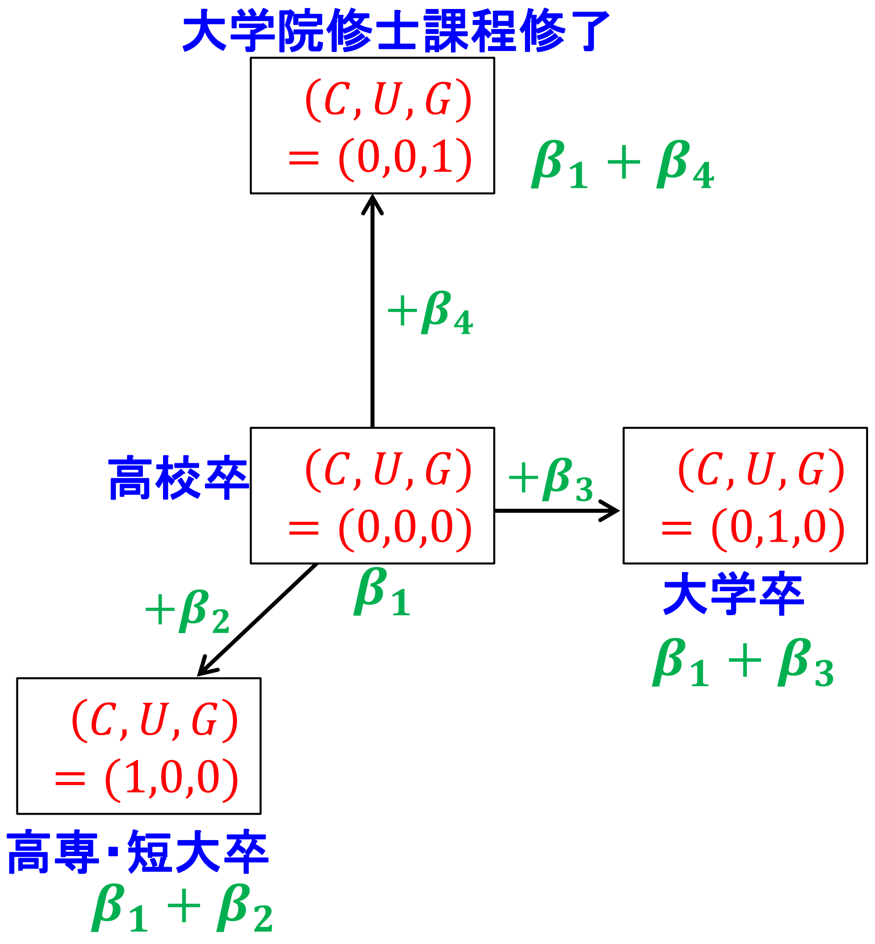
(p156.1) [C10-1]問4. ダミー変数・単回帰係数の性質

[1]:BCランク
[2]:BCランク

小問[1]

初任給: $y=\beta_1+\beta_2 C+\beta_3 U+\beta_4 G+u$





(p156.2) [C10-1]問4. ダミー変数・単回帰係数の性質

[1]:BCランク
[2]:BCランク

小問[1]

I

・高校卒ダミー変数Hを導入した場合、

$$
\begin{aligned}
(H, C, U, G) = \text{高校卒}(1,0,0,0),\\
\text{高専・短大卒}(0,1,0,0),\\
\text{大卒}(0,0,1,0),\\
\text{大学院修士課程修了}(0,0,0,1)
\end{aligned}
$$

となり、$H + C + U + G = 1 \Rightarrow H = 1 - (C + U + G)$
C,U,Gが決まれば、Hが決まり、独立な変数ではなくなる。
多重共線性の問題が生じてしまう。

I: 正しくない

$$
\begin{aligned}
y &= \gamma_1 + \gamma_2 H + \gamma_3 C + \gamma_4 U + \gamma_5 G + v\\
&= \gamma_1 + \gamma_2 (1 - (C + U + G)) + \gamma_3 C + \gamma_4 U + \gamma_5 G + v
\end{aligned}
$$

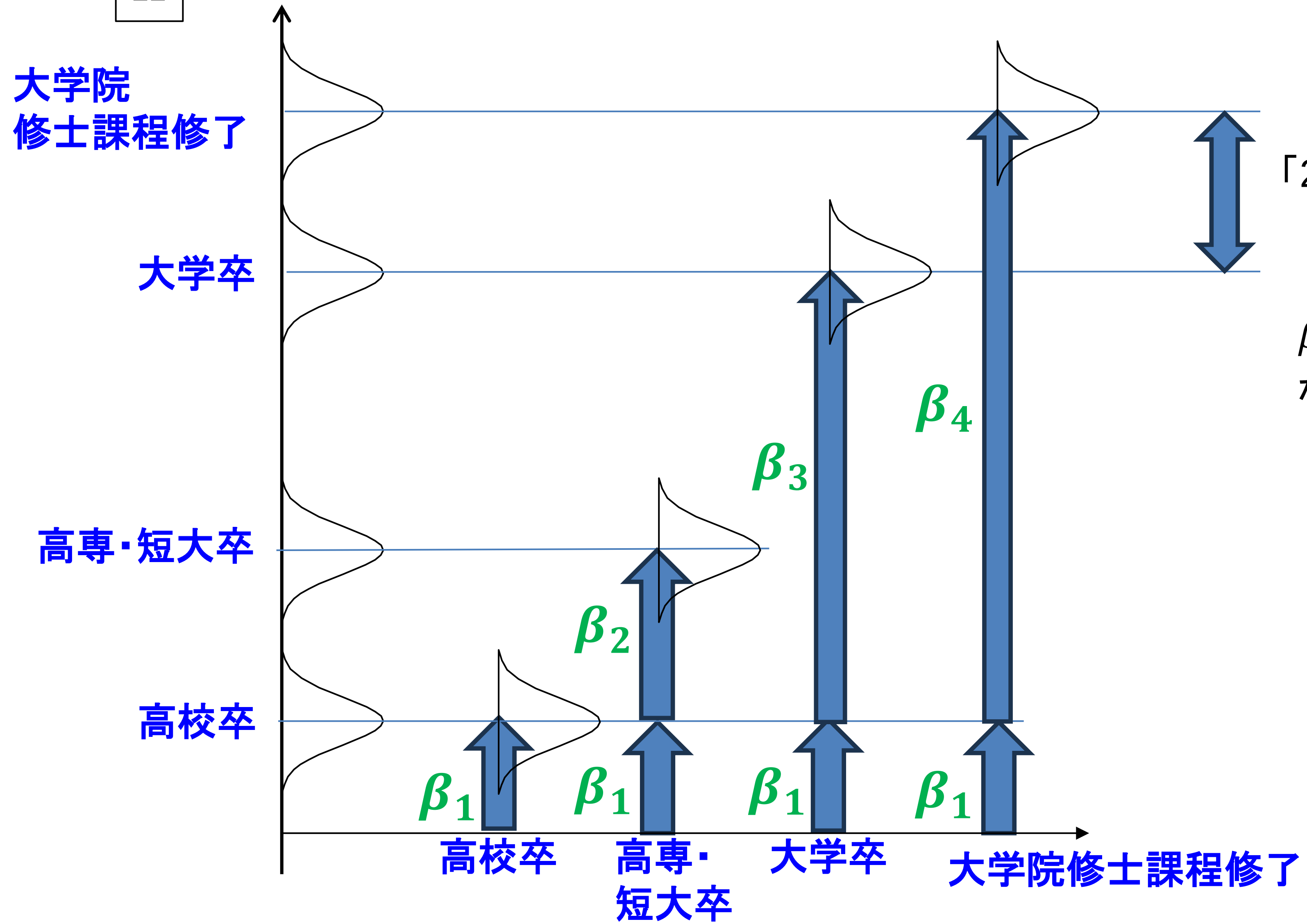
$\Downarrow$

$$
y = \beta_1 + \beta_2 C + \beta_3 U + \beta_4 G + u
$$

(p156.3) [C10-1]問4. ダミー変数・単回帰係数の性質

[1]:BCランク
[2]:BCランク

小問[1] II



「2.73万円高い」は、
正しい？

$\beta_1 = 16.653$
$\beta_2 = 2.255$
$\beta_3 = 4.450$
$\beta_4 = 7.180$

$\beta_4 - \beta_3 = 7.18 - 4.45 = 2.73$
なので、正しい

II: 正しい

(p156.4) [C10-1]問4. ダミー変数・単回帰係数の性質

[1]:BCランク
[2]:BCランク

小問[1]

III

1. 観測数:n=16 (学歴:4種類×業種:4)
全自由度=16-1=15
回帰の自由度=3 (説明変数は、(C,U,G)の3個なので)
残差の自由度=15-3=12

自由度13ではなく、自由度12のt分布が使われている

III: 正しくない

回帰分析の自由度:
データ総数:nの時、
全自由度$\boldsymbol{\phi}_T = n - 1$
回帰の自由度: (単回帰では) $\boldsymbol{\phi}_R = 1$
（説明変数がp個の重回帰では、$\boldsymbol{\phi}_R = p$)

残差の自由度:$\boldsymbol{\phi}_e = \boldsymbol{\phi}_T - \boldsymbol{\phi}_R$

(公式) $t = \dfrac{(a)\text{推定値} - \beta_1}{(b)\text{標準誤差}}$ ～自由度ϕ_eのt分布

I:	正しくない
II:	正しい
III:	正しくない

[1](答)②

(p156.5) [C10-1]問4.ダミー変数・単回帰係数の性質

[1]:BCランク
[2]:BCランク

小問[2]

小問[1]:重回帰分析 ⇒ 小問[2]:単回帰分析(x: 教育年数)

I

得られた回帰直線: $y = 2.323 + 1.187x$

$2.323 + 1.187x = y$ の時、
$2.323 + 1.187(x + 1) = y + 1.187$
となるので、正しい

I: 正しい

II

決定係数 =0.898 であり、
自由度調整済決定係数 =0.891 ⇒違いは計算の丸め誤差？

今の例では、観測数=n=16
全自由度 $\phi_T = n - 1 = 16 - 1 = 15$
回帰の自由度 $\phi_R = 1$
残差の自由度 $\phi_e = 15 - 1 = 14$

決定係数=0.898の時、

$$\frac{S_e}{S_T} = 1 - 0.898 = 0.102$$

決定係数:
$= 1 - 0.102 \quad = 0.898$

自由度調整済決定係数:

$$= 1 - 0.102 \times \left(\frac{15}{14}\right) = 0.891$$

II: 正しくない

●決定係数(寄与率):

$$R^2 = 1 - \frac{S_e}{S_T}$$

●自由度調整済決定係数:

$$R^{*2} = 1 - \frac{\frac{S_e}{\phi_e}}{\frac{S_T}{\phi_T}} = 1 - \frac{S_e}{S_T}\frac{\phi_T}{\phi_e}$$

S_T: 全平方和
S_e: 残差平方和
ϕ_T: 全自由度
ϕ_e: 残差の自由度

(p156.6) [C10-1]問4.ダミー変数・単回帰係数の性質

[1]:BCランク
[2]:BCランク

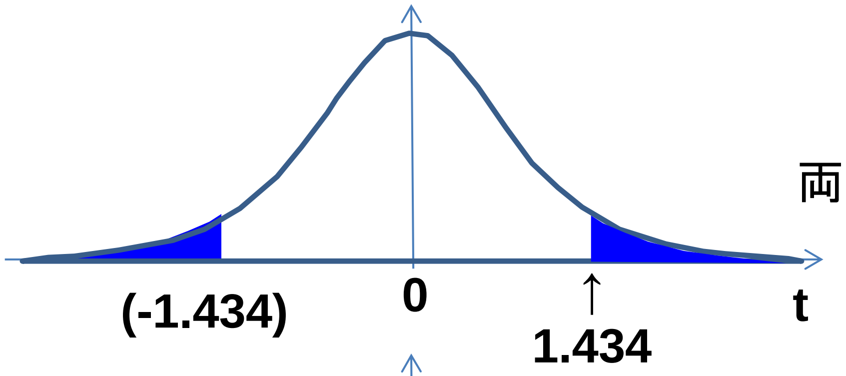
小問[2] III

回帰直線: $y = \alpha + \beta x$

両側検定:
帰無仮説 H_0: $\alpha = 0$
対立仮説 H_1: $\alpha \neq 0$

片側検定:
H_0: $\alpha = 0$
H_1: $\alpha > 0$

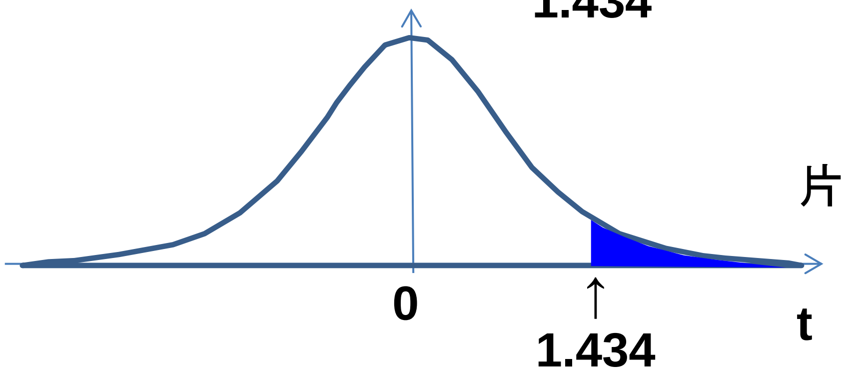
P値は同じ？



両側検定

P値は
半分になる

片側検定



III: 正しくない

I: 正しい
II: 正しくない
III: 正しくない

[2](答)①

済