予測を確かめ、信用できる範囲を考える

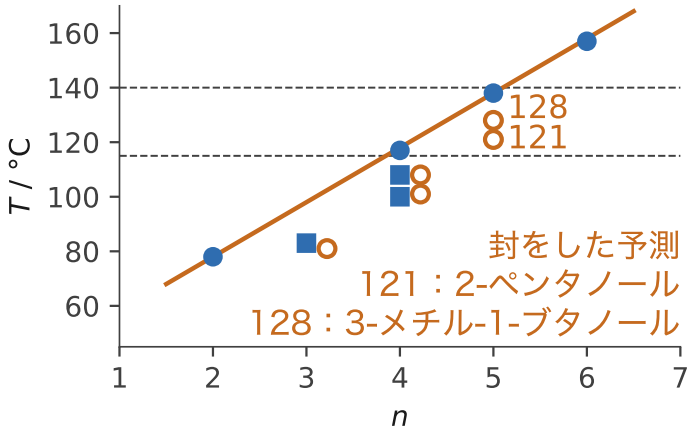
分子設計 第 5 回 ｜ 教科書 第 5 章

第 5 話　封を開ける

予測がよくなったことと、信用できることはどう確かめるか

これまでの経過

第 4 回　構造と沸点の傾向を、簡単な式にまとめられるか



直線：M0　●■：資料値
○：M1 の予測　破線：要求の範囲

今日できるようになったこと

- ○ 直線の式 M0 で予測し、残差（資料値 − 予測値）を求められる
  2-プロパノール：予測 98 ℃、残差 −15 ℃
- ○ 残差のそろい方から補正の項を入れ、係数を SSE で選べる
  補正の係数は二級・分岐の 3 分子から。c_s は 15・17・20 から 17（16 も同じ SSE）
- ○ 係数を決めた分子への当てはまりと、未知の分子での予測を分けて言える

まだ分からないこと

- ？ M1 の予測は、使っていない分子でも合うか → 第 5 回
- ？ 最近傍と M1 の予測の、どちらを信用するか → 第 5 回
  記録：2-ペンタノール 138・100・119・121 ℃、3-メチル-1-ブタノール 128 ℃

M0（炭素数だけの式）$T = 20n + 38$　|　**M1**（補正式）$T = 20n + 38 - 17s - 10b$

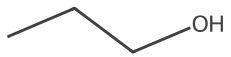
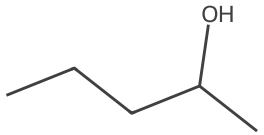
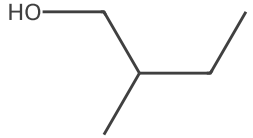
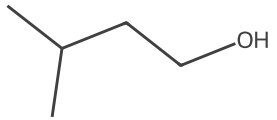
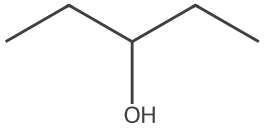
依頼：沸点 115〜140 ℃　|　w_{25} 4.0 以上　|　C5〜C6 の一級・二級

前回の確認 1　c_b は 2-メチル-1-プロパノール 1 分子にぴったり合わせた（108 ℃）。その分子で誤差 0 になったことだけで、未知の分子への予測がよいと言えるか　（言える／言えない）

問い 1　予測がよくなったことと、信用できることはどう確かめるか



工程担当者　「資料が届いたよ。その予測、本当に信じていいの？」

				
1-プロパノール	**2-ペンタノール**	**2-メチル-1-ブタノール**	**3-メチル-1-ブタノール**	**3-ペンタノール**
$(3, 0, 0)$	$(5, 1, 0)$	$(5, 0, 1)$	$(5, 0, 1)$	$(5, 1, 0)$
資料値 ？	**資料値 ？**	**資料値 ？**	**資料値 ？**	**資料値 ？**

式にも参照分子にも使っていない 5 分子で確かめる

今日の流れ

教科書 §5.1〜5.7

依頼：1-ペンタノールを改良し、候補と次の実験を提案する（第 1〜6 回）

データを整える	構造を表す	予測する		確かめる	候補と次の実験
第 1 回 ✓ データ	第 2 回 ✓ 記述子	第 3 回 ✓ 似た分子	第 4 回 ✓ 予測式	**第 5 回 検証**	**第 6 回** 提案
				▲ 今日	

1. **問い 1 ①** 予測がよくなったことを、どう確かめるか → 例題・基本演習 A・B・C
2. **問い 1 ②** その予測を、どこまで信用できるか → 例題・応用演習 1・2

そのあと：問い 1 の答え → 実際のケモインフォでは → 自分の実験では

当てはまりと、予測を分ける

教科書 §5.1

問い 1 ①　どのデータで確かめるか

直鎖一級の 4 分子

第 1 回

エタノール・1-ブタノール・
1-ペンタノール・1-ヘキサノール

両端の 2 分子から（傾きは丸めた）M0 の直線を決めた。1-ペンタノールは最近傍の参照分子にも使った

二級・分岐の 3 分子

第 3・4 回

2-プロパノール・2-ブタノール・
2-メチル-1-プロパノール

c_s・c_b を決めた（うち 2 つは最近傍の参照分子）

→

確かめる 5 分子

第 5 回（今日）

1-プロパノール・2-/3-ペンタノール・
2-/3-メチル-1-ブタノール

式にも参照分子にも使っていない。予測を記録してから資料値を見る

- **当てはまり**：式や参照分子を決めるときに使ったデータに、予測が合うこと
- **予測の確認**：使っていないデータで、予測を記録してから資料値と比べること

基本演習 A：予測を記録する

問い 1 ①　確かめる前に、何をするか

M0：$T = 20n + 38$　　M1：$T = 20n + 38 - 17s - 10b$

手順 1　1-プロパノールの M0・M1 の予測を手で求める

手順 2　残りの空欄も求める

手順 3　アプリで自分の値と照らし合わせ、「予測を記録する」を押す

問　M1 の予測が同じになった分子の組は、(n, s, b) がどうなっているか

分子	(n, s, b)	M0	M1
1-プロパノール	(3, 0, 0)		
2-ペンタノール	(5, 1, 0)		121
2-メチル-1-ブタノール	(5, 0, 1)		
3-メチル-1-ブタノール	(5, 0, 1)		128
3-ペンタノール	(5, 1, 0)		

121・128：前回封をした記録。単位は ℃

基本演習 A：アプリで記録する

分子	(n, s, b)	M0	M1	資料値	M0 の絶対誤差	M1 の絶対誤差
1-プロパノール	$(3, 0, 0)$					
2-ペンタノール	$(5, 1, 0)$		121			
2-メチル-1-ブタノール	$(5, 0, 1)$					
3-メチル-1-ブタノール	$(5, 0, 1)$		128			
3-ペンタノール	$(5, 1, 0)$					

空欄に M0・M1 の予測を書き込む

終わったら：前の枚の問へ

基本演習 A の答え

記録した予測（M0、M1）

① **1-プロパノール** 98、98 ℃

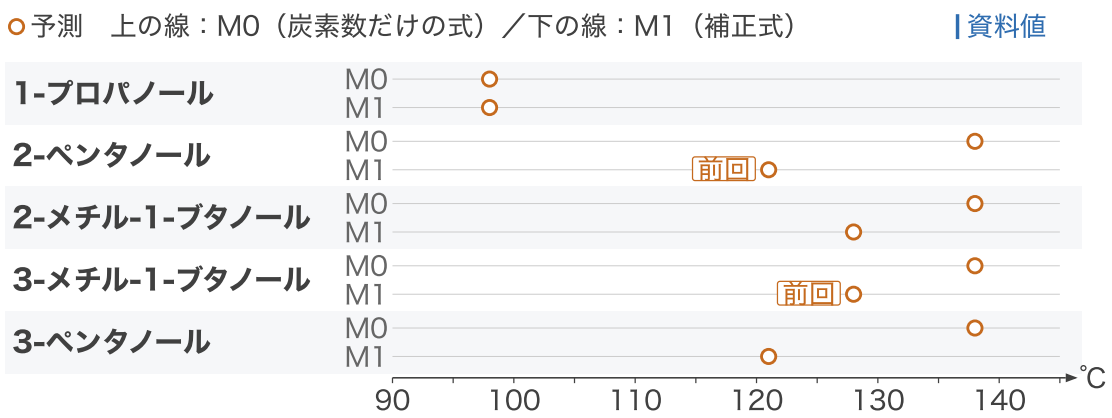
② **2-ペンタノール** 138、121 ℃

③ **2-メチル-1-ブタノール** 138、128 ℃

④ **3-メチル-1-ブタノール** 138、128 ℃

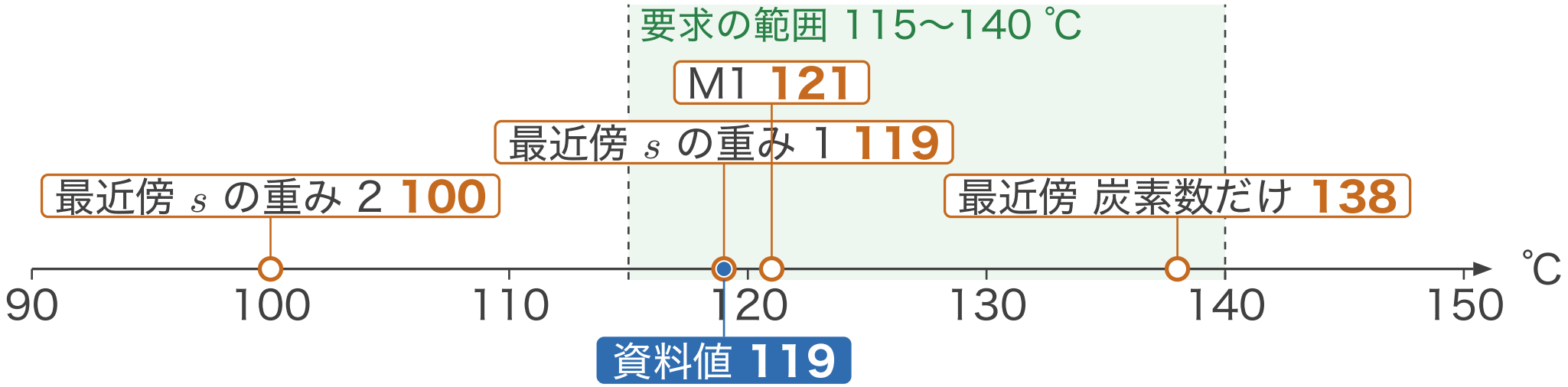
⑤ **3-ペンタノール** 138、121 ℃

答え　M1 の予測が同じ組は、(n, s, b) が同じ



封を開ける

問い 1 ①　1 分子の一致で、決めてよいか



重み 1 がぴったり。でも、これは 1 分子だけ → 5 分子で比べる

2-ペンタノール。白抜き：封をした予測　青：資料値

予測のずれを、1 つの数にまとめる：MAE

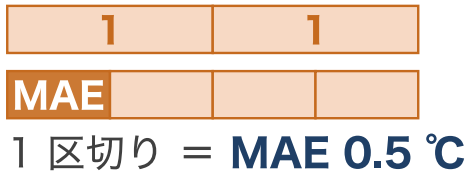
教科書 §5.2

問い 1 ①　複数の分子のずれを、どうまとめるか

$$\mathrm{MAE} = \frac{1}{N}\left(|e_1| + |e_2| + \cdots + |e_N|\right)$$

N 分子の数　$e_1, e_2, \ldots$ 分子ごとの残差（資料値 − 予測値。第 4 回）　$|e|$ 絶対誤差

分子	資料値	M0 予測	絶対誤差	棒の長さ＝絶対誤差
エタノール	78	78	$\lvert 78-78\rvert = 0$	
1-ブタノール	117	118	$\lvert 117-118\rvert = 1$	
1-ペンタノール	138	138	$\lvert 138-138\rvert = 0$	
1-ヘキサノール	157	158	$\lvert 157-158\rvert = 1$	



1 区切り ＝ **MAE 0.5 ℃**

まとめ方の練習（M0 と直鎖一級 4 分子）。分子が違う MAE とは比べない

例題：係数を決めた 3 分子の MAE

問い 1 ①　複数の分子のずれを、どうまとめるか

問題　M1 の予測は 2-プロパノール 81 ℃、2-ブタノール 101 ℃、2-メチル-1-プロパノール 108 ℃。資料値は 83・100・108 ℃。MAE を求めてください

解き方

① 絶対誤差

$$|83-81| = 2$$
$$|100-101| = 1$$
$$|108-108| = 0$$

② 平均

$$(2+1+0)/3 = 1.0\ ℃$$

③ 読む　係数を決めた 3 分子なので当てはまり

答え　MAE = 1.0 ℃（当てはまりの MAE）

基本演習 B：M0 と M1 を MAE で比べる

問い 1 ①　予測はよくなったか

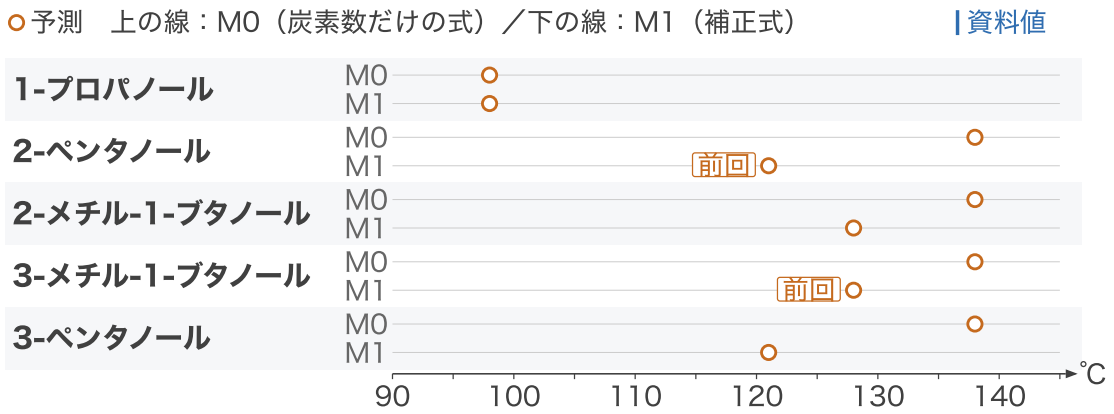
手順 1　2-ペンタノール（資料値 119、M0 138、M1 121 ℃）の絶対誤差を手で求める

手順 2　予想：使っていない 5 分子の M1 の MAE は、係数を決めた 3 分子の 1.0 ℃（例題）より　（大きい／小さい）

手順 3　アプリで資料値を表示し、M1 の MAE を手で求める

手順 4　アプリで「絶対誤差と MAE を表示」を押し、M0 と比べる

問　M0 → M1 で MAE は　（小さくなった／大きくなった）



基本演習 B：アプリで確かめる

		式		最近傍（第 3 回）			
分子	(n, s, b)	M0	M1	炭素数だけ	$\alpha = 2$	$\alpha = 1$	資料値
1-プロパノール	$(3, 0, 0)$	98	98	104	123	115.3	97
2-ペンタノール	$(5, 1, 0)$	138	121	138	100	119	119
2-メチル-1-ブタノール	$(5, 0, 1)$	138	128	138	123	123	128
3-メチル-1-ブタノール	$(5, 0, 1)$	138	128	138	123	123	132
3-ペンタノール	$(5, 1, 0)$	138	121	138	100	119	116

M0・M1 と資料値の列から、絶対誤差と MAE を求める

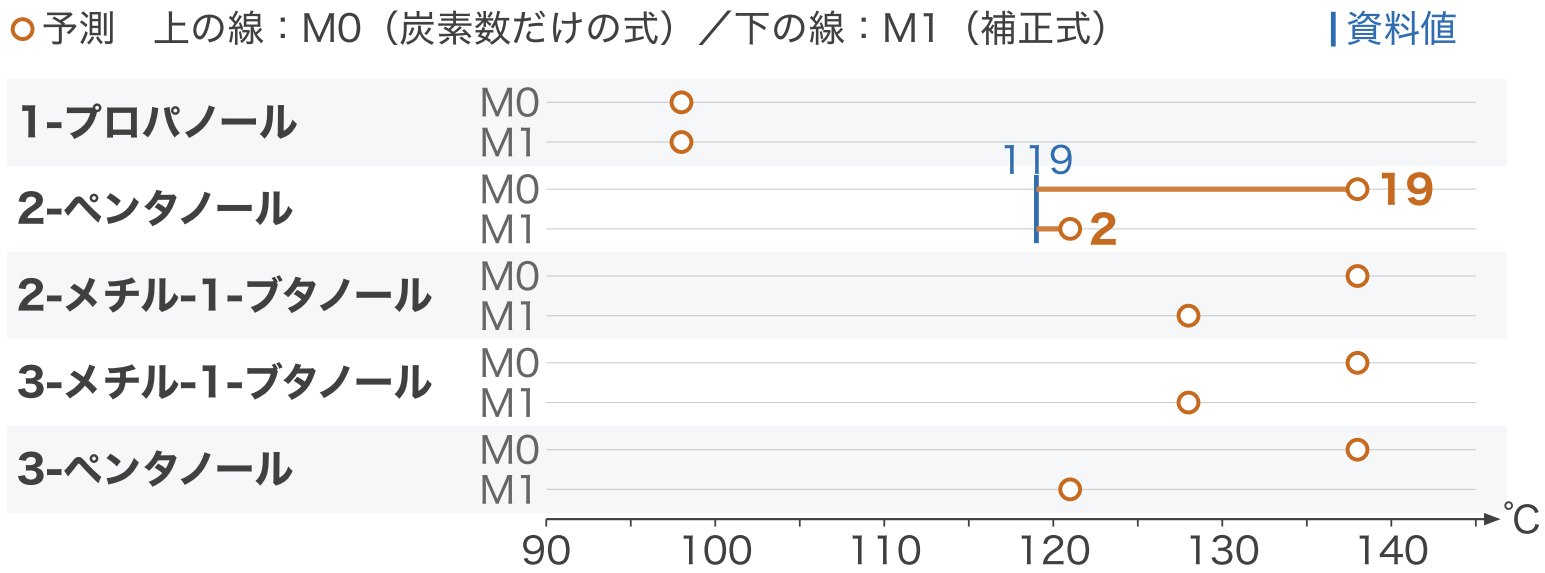
終わったら：前の枚の問へ

基本演習 B の答え（1）

手順 1　2-ペンタノールの絶対誤差

考え方　資料値 119 ℃。M0：$|119 - 138| = 19$、M1：$|119 - 121| = 2$

答え　M0 は 19 ℃、M1 は 2 ℃



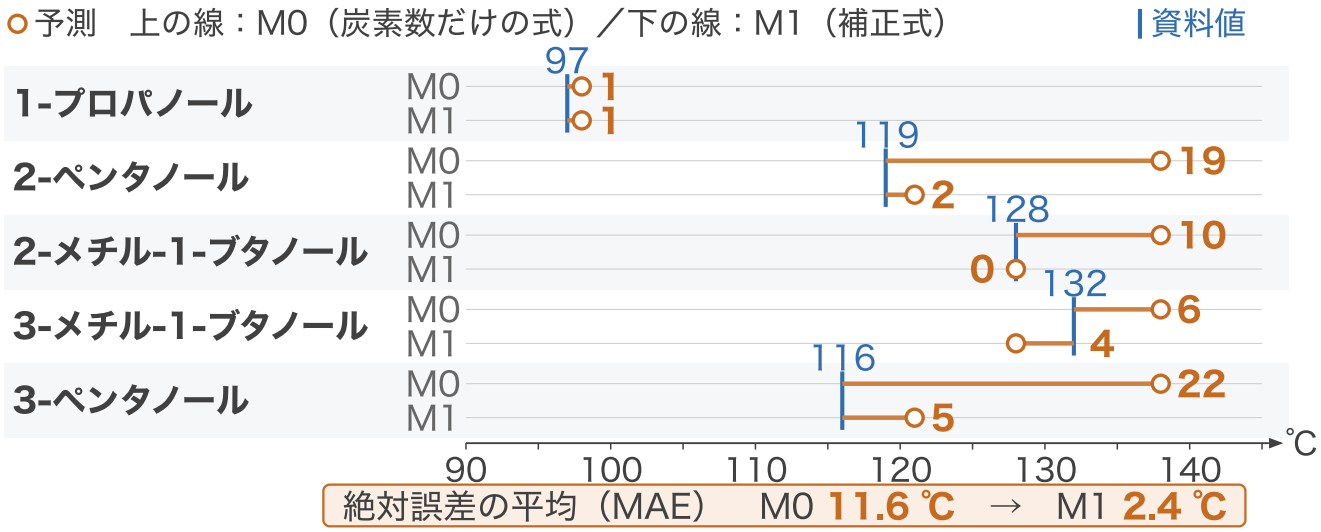
基本演習 B の答え（2）

手順 2・3、問　5 分子の MAE

① M0　絶対誤差の和 58 ℃。$58/5 = 11.6$ ℃

② M1　絶対誤差の和 12 ℃。$12/5 = 2.4$ ℃（例題の 1.0 ℃ より大きい）

答え　小さくなった



要点：記録してから確かめた。この 5 分子では、M1 は M0 より予測がよい

基本演習 C：最近傍とも比べる

問い 1 ①　1 分子の一致で、決めてよいか

2-ペンタノール：資料値 119 ℃。最近傍の予測は炭素数だけ 138、s の重み $\alpha = 2$ で 100、$\alpha = 1$ で 119 ℃（第 3 回）

最近傍の参照分子（第 3 回）　**1-ペンタノール** $(5, 0, 0)$ **138 ℃**　**2-ブタノール** $(4, 1, 0)$ **100 ℃**　**2-メチル-1-プロパノール** $(4, 0, 1)$ **108 ℃**

手順 1　3 つの絶対誤差を手で求める（M1 は 2 ℃）

手順 2　アプリで 5 分子の MAE を確かめる

問　$\alpha = 1$ は 2-ペンタノールで誤差 0。5 分子の MAE が最小の方法は（最近傍（α = 1）／M1）

基本演習 C：アプリで確かめる

		式		最近傍（第 3 回）			
分子	(n, s, b)	M0	M1	炭素数だけ	$\alpha = 2$	$\alpha = 1$	資料値
1-プロパノール	$(3, 0, 0)$	98	98	104	123	115.3	97
2-ペンタノール	$(5, 1, 0)$	138	121	138	100	119	119
2-メチル-1-ブタノール	$(5, 0, 1)$	138	128	138	123	123	128
3-メチル-1-ブタノール	$(5, 0, 1)$	138	128	138	123	123	132
3-ペンタノール	$(5, 1, 0)$	138	121	138	100	119	116

最近傍（参照分子 3 つ）の 5 分子の予測。第 3 回と同じ手順で距離から求めた。MAE は答えの枚に

終わったら：前の枚の問へ

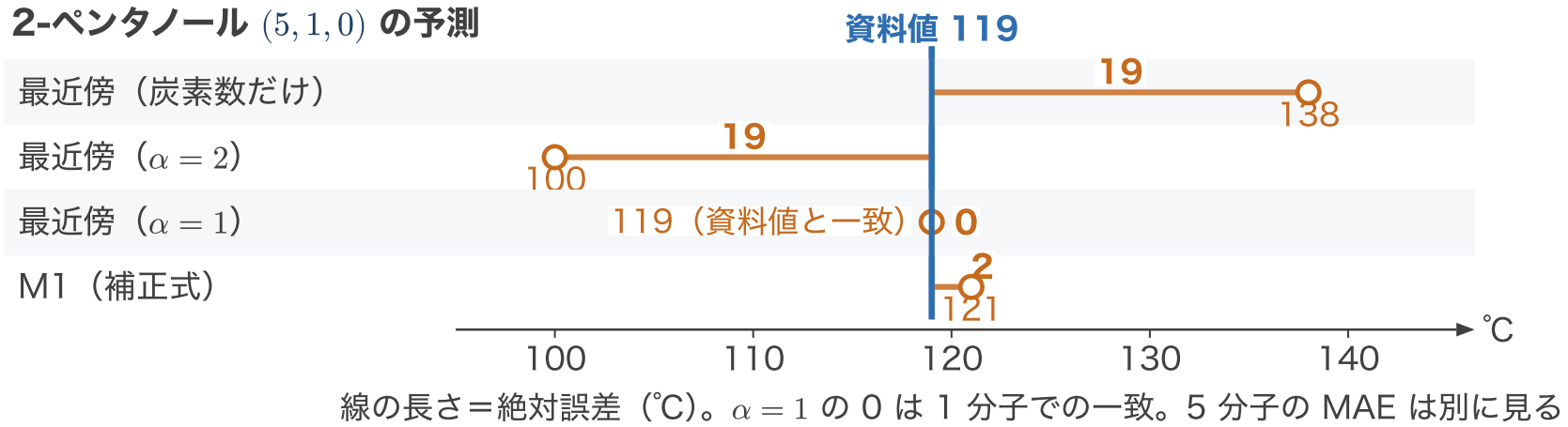
基本演習 C の答え

5 分子の MAE が最も小さい方法は

① 2-ペンタノール 炭素数だけ 19、$\alpha = 2$ で 19、$\alpha = 1$ で 0 ℃（M1 は 2 ℃）

② 5 分子の MAE 炭素数だけ 12.8、$\alpha = 2$ で 15.0、$\alpha = 1$ で 7.1、M1 2.4 ℃

答え　この 5 分子では M1



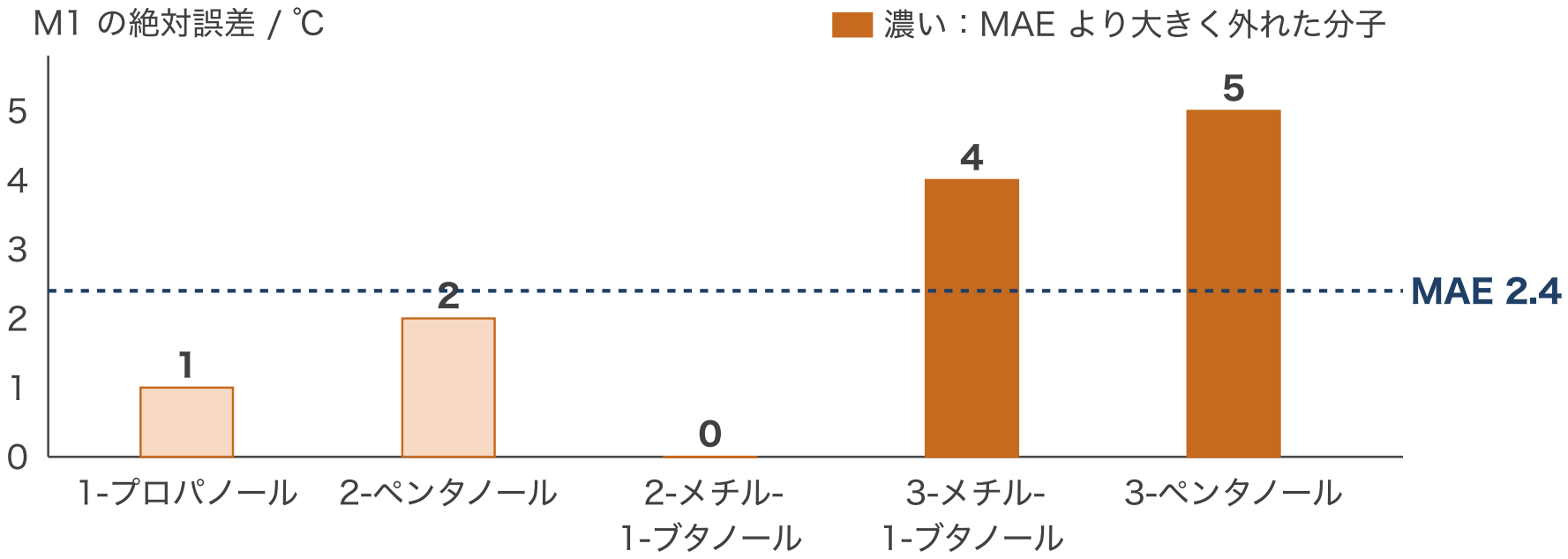
線の長さ＝絶対誤差（℃）。$\alpha = 1$ の 0 は 1 分子での一致。5 分子の MAE は別に見る

要点：1 分子の一致では決められない。方法どうしは同じ分子で比べる

MAE 2.4 ℃ から、何が言えるか

教科書 §5.2

問い 1 ② その予測を、どこまで信用できるか：MAE は誤差の上限か



- MAE は平均。「どの分子も 2.4 ℃ 以内」とは読めない（4 ℃・5 ℃ の分子がある）

見たデータでは確かめられない：データ漏洩

教科書 §5.3

問い 1 ②　資料値を見て直したら、どうなるか

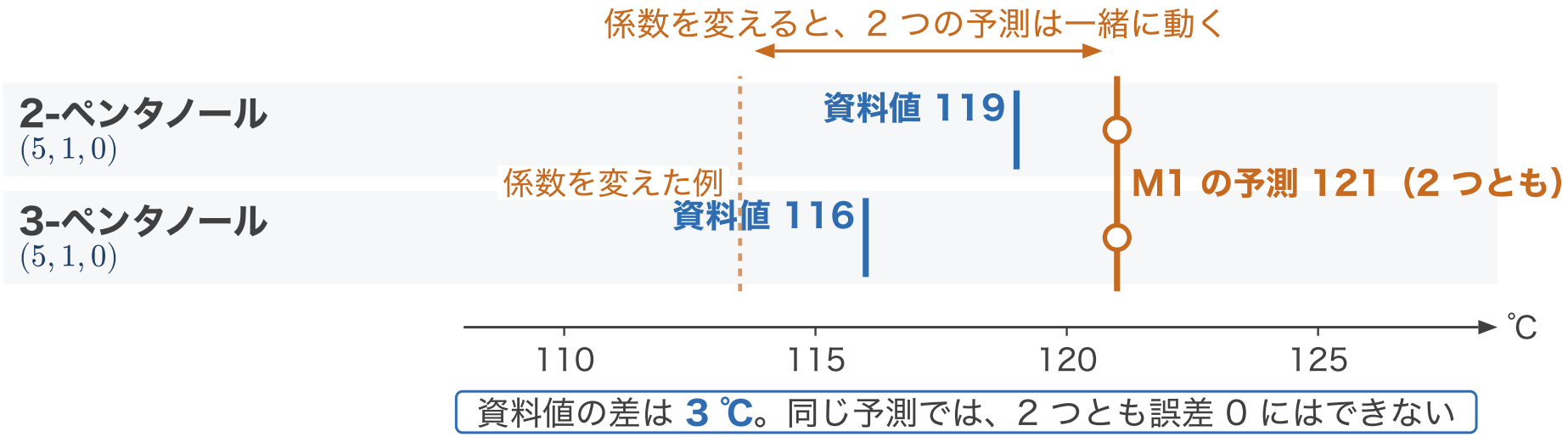


- **データ漏洩**：確かめるためのデータを見て、式や重み、係数を決めたり直したりすること。確かめたつもりの成績が、実際より良く見える

同じ記述子なのに、沸点が違う

教科書 §5.4

問い 1 ② 係数を変えれば、両方に合わせられるか

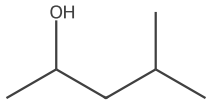


2-ペンタノールと 3-ペンタノールは同じ (5, 1, 0)。係数を変えれば、2 つとも誤差 0 にできるか（できる／できない）

どこまで使ってよいか：適用範囲

教科書 §5.5

問い 1 ②　式に入れられれば、信用できるか

(s,b) ＼ n	2	3	4	5	6	…	$n=9$（代入はできる）
$(0,0)$ 直鎖一級	●	●	●	●	●		範囲外
$(1,0)$ 二級		●	●	●			範囲外
$(0,1)$ 分岐			●	●			範囲外
$(1,1)$ 二級＋分岐			4-メチル-2-ペンタノール →		?		範囲外

● 式を決めるのに使った分子　● 確かめた分子（使っていない 5 分子）　? まだ資料値がない

- **適用範囲**：資料値のある分子（式を決めた分子と、確かめた分子）が覆う範囲。列ごとの範囲と、組み合わせの両方を見る

理解確認（個人・3 分）

MAE＝絶対誤差の平均。予測 100・102 ℃、資料値はどちらも 100 ℃（架空）

1. 絶対誤差 2 つと MAE を求める
2. 個々の誤差は、すべて MAE 以下か　（すべて以下／そうではない）
3. 資料値を見て係数を変えた。変えた後の予測のよさは、何のデータで確かめるか



理解確認の答え

問 1・2　絶対誤差と MAE。個々の誤差は、すべて MAE 以下か

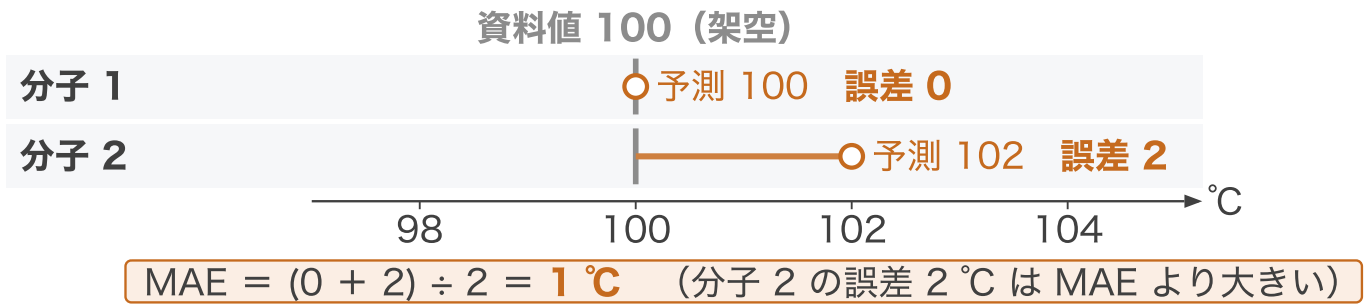
考え方　$|100-100|=0$、$|100-102|=2$。MAE $=(0+2)/2=1$ ℃

答え　1：0・2 ℃、MAE 1 ℃　2：そうではない（2 ℃ の分子は MAE より大きい）

問 3　係数を変えた後のよさは、何のデータで確かめるか

よくある答え　「同じデータでもう一度」は ×（データ漏洩）。
「もっと多くのデータで」は、係数を直すのに使っていなければ ○

答え　係数を直すのに使っていない、別のデータ



例題：主張の言える範囲を確かめる

問い 1 ②　主張をどう確かめるか

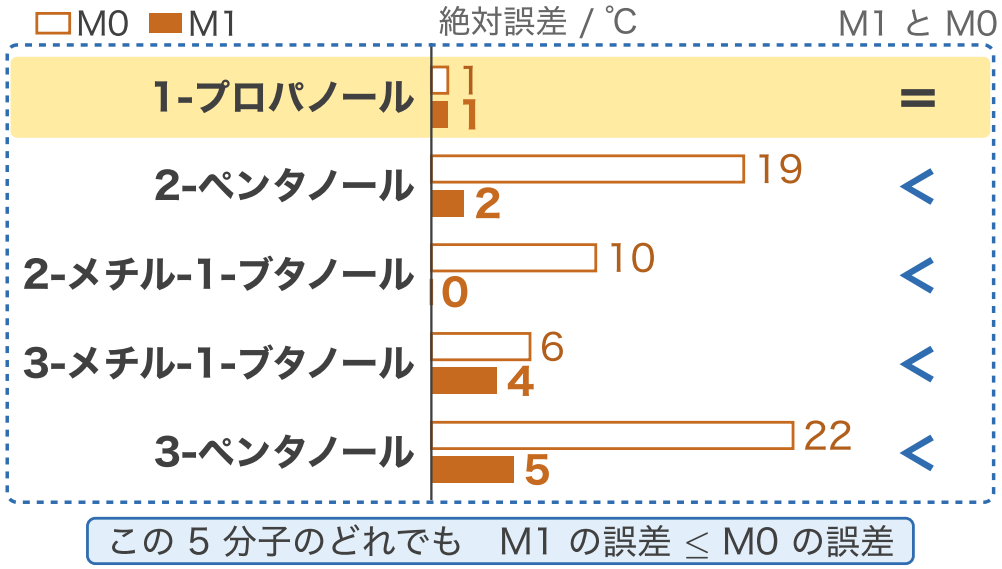
> **問題**　「5 分子のどれでも、M1 は M0 より誤差が小さかった」と言えるか。右の図から判定してください

解き方

① 範囲　主張は「どれでも」
→ 5 行すべてを見る

② 確かめる　1-プロパノールは、
M0・M1 とも誤差 1 ℃（同じ）

③ 直す　「この 5 分子のどれでも、M1 の誤差は M0 以下」なら言える



答え　言えない（1 分子は同じ）。言い方を「以下」に直せば言える

応用演習 1：結果から言えること

問い 1 ②　主張をどう確かめるか

5 分子の M1 の絶対誤差：1・2・0・4・5 ℃（MAE 2.4 ℃。M0 の MAE は 11.6 ℃）

手順 1　　主張が何のデータ、どの範囲の話かに印をつける（「この 5 分子」「必ず」など）

手順 2　　根拠を確かめ、言える／言えないを選ぶ。理由を 1 行で

手順 3　　主張 2 は、言える文に直す

主張 1　「この 5 分子では、M1 は M0 より予測がよかった」　（言える／言えない）

主張 2　「新しい分子でも、M1 なら必ず ±2.4 ℃ 以内に入る」　（言える／言えない）

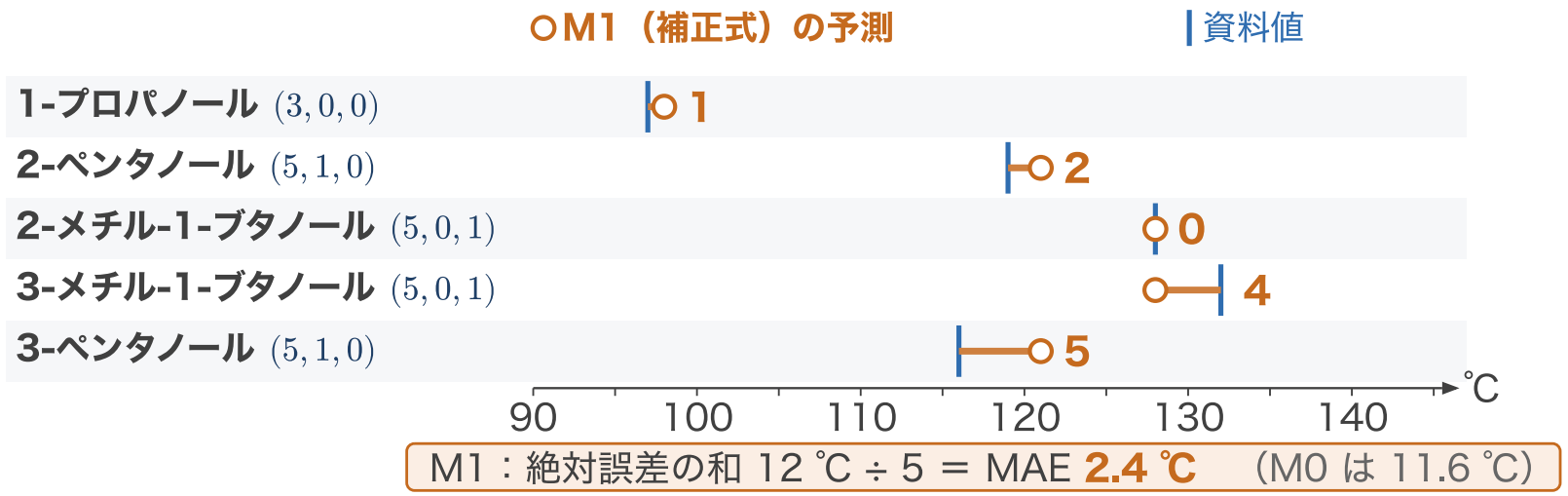
主張 3　「今日の 5 分子の資料値を見て係数を直しても、同じ 5 分子でよさを確かめられる」（言える／言えない）

応用演習 1 の答え（1）

主張 1　この 5 分子では、M1 は M0 より予測がよかった

根拠　使っていない 5 分子で、記録してから確かめた。MAE 11.6 → 2.4 ℃

答え　言える（「この 5 分子では」の範囲で）



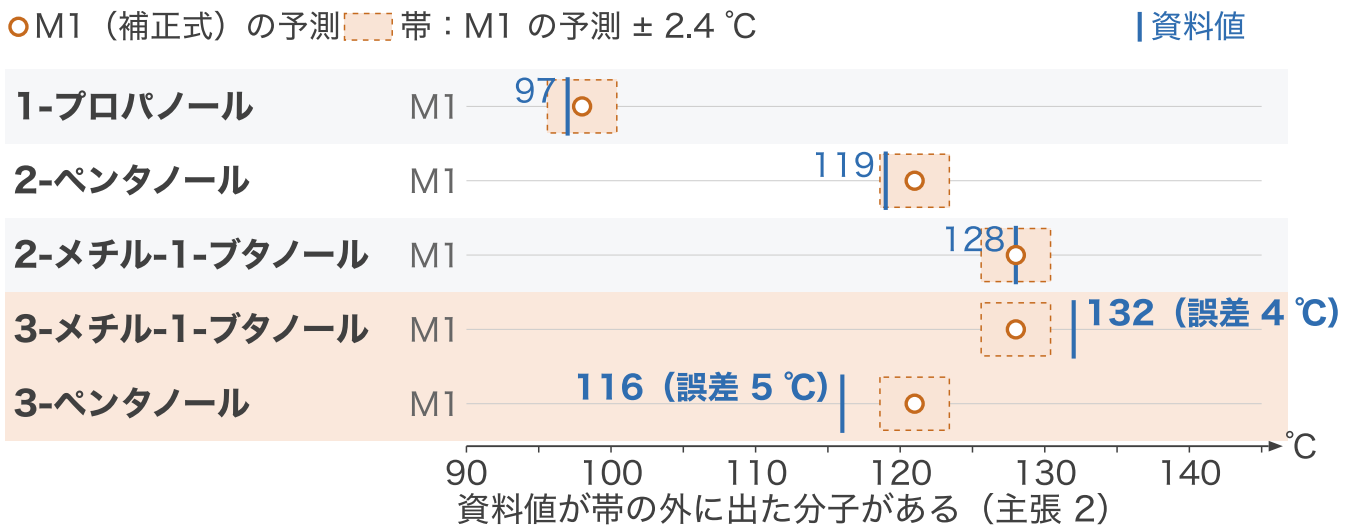
応用演習 1 の答え（2）

主張 2　新しい分子でも、M1 なら必ず ±2.4 ℃ 以内

考え方　MAE は平均で、上限ではない

直した文　「この 5 分子では、M1 の絶対誤差は平均 2.4 ℃、最大 5 ℃ だった」

答え　言えない



応用演習 1 の答え（3）

主張 3　資料値を見て係数を直した後も、同じ 5 分子で確かめられる

考え方　直した時点で、5 分子は係数を直すのに使ったデータ

答え　言えない。別の使っていないデータで確かめる



応用演習 2：組み合わせと係数の数

問い 1 ②　主張をどう確かめるか

資料値のある分子：n = 2〜6、(s, b) は (0, 0)、(1, 0)、(0, 1) だけ

手順 1　　主張が何のデータ、どの範囲の話かに印をつける

手順 2　　根拠を確かめ、言える／言えないを選ぶ。理由を 1 行で

手順 3　　主張 4 は、言える文に直す

主張 4　「n・s・b がそれぞれ確かめた範囲なら、(6, 1, 1) の予測も確かめたことになる」　（言える／言えない）

主張 5　「分子ごとに別の補正係数を付ければ、第 4 回までの 7 分子（直鎖一級 4＋二級・分岐 3）の SSE は 0 にできる。だから予測もよくなる」　（言える／言えない）

応用演習 2 の答え（1）

主張 4　各列が範囲内なら、(6, 1, 1) も確かめたことになる

考え方　資料値のある分子に、二級＋分岐の組み合わせは 1 つもない

直した文　「(6, 1, 1) は式で値は出るが、二級＋分岐の組み合わせは確かめていない」

答え　言えない。式に入れられることと、信用できることは別

(s,b) ＼ n	2	3	4	5	6	
$(0,0)$ 直鎖一級	●	●	●	●	●	資料値のある n：2〜6 ✓
$(1,0)$ 二級		●	●	●		s：0 か 1 ✓
$(0,1)$ 分岐			●	●		b：0 か 1 ✓
$(1,1)$ 二級＋分岐					?	**組み合わせ：実績なし**

$(6,1,1)$ は各列の範囲内。ただし資料値のある分子に、$s=b=1$ は 1 つもない

応用演習 2 の答え（2）

主張 5　分子ごとに係数を付ければ SSE は 0。だから予測もよくなる

考え方　使ったデータに合わせすぎた式（**過学習**）。どの資料値も再現できるが、新しい分子にどの係数を使うか決まらない

答え　言えない。当てはまりが完全でも、予測がよいとは限らない

M1：補正の係数 2 つ（c_s・c_b）

第 4 回までの 7 分子に、近いがぴったりではない

$-17s$　$-10b$

新しい分子 → 同じ式に (n, s, b) を入れれば予測できる

分子ごとに補正の係数（7 つ）

第 4 回までの 7 分子には、ぴったり合う（SSE = 0）

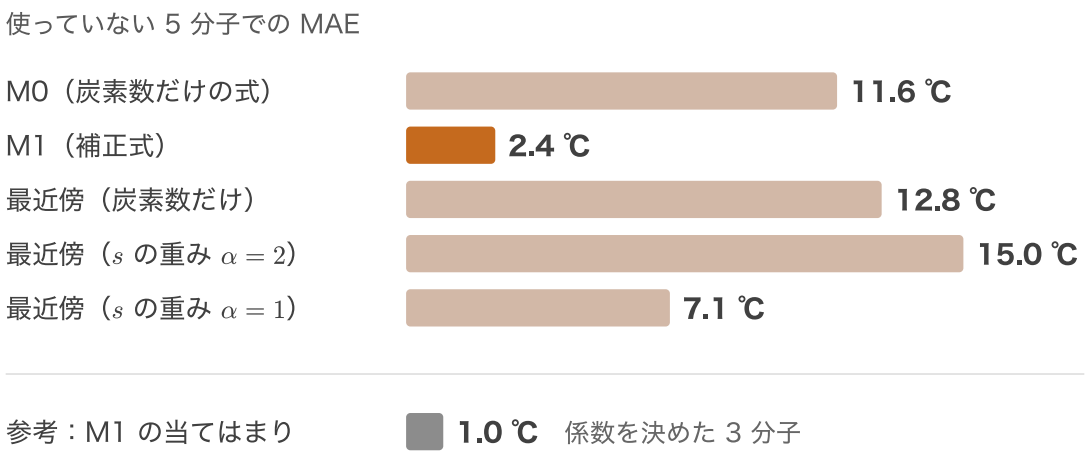
c_1　c_2　c_3　c_4　c_5　c_6　c_7

新しい分子 → どの係数を使う？ **決まらない**

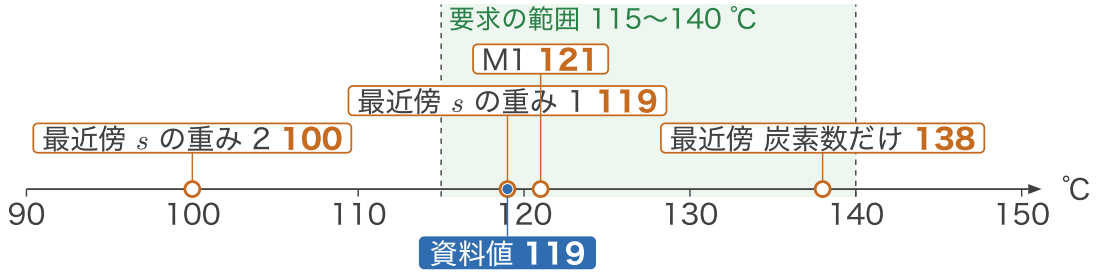
要点：言えるのは確かめた範囲まで。
平均は上限でなく、見たデータも合わせすぎも、予測のよさを示さない

問い 1 の答え 使っていないデータで確かめ、言える範囲を限る

問い 1：予測がよくなったことと、信用できることはどう確かめるか



使っていない 5 分子の MAE（灰：係数を決めた 3 分子）



2-ペンタノールの 4 つの予測と資料値

- よくなった：記録してから、使っていない 5 分子で MAE 11.6 → 2.4 ℃
- 言える範囲：この 5 分子の平均で、上限ではない。$s = b = 1$ の組み合わせは未確認

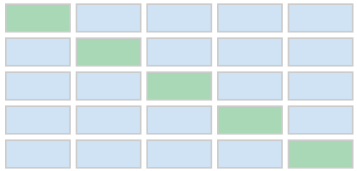
実際のケモインフォでは：使っていないデータで確かめる

学習用と検証用に分ける



青で式を決め、緑で確かめる。
緑を見て式を直さない
（データ漏洩を防ぐ）

交差検証



確かめる部分を入れ替えて、
何度も確かめる

過学習と適用範囲

列や係数を増やすと、学習用には
よく合うが、検証用で外れる
ことがある（過学習）。
学習用のデータから遠い分子では、
予測を信用しにくい（適用範囲）

この授業：使っていない 5 分子で、1 回確かめた

実際の研究：分け方を変えて何度も確かめ、範囲も示す

決めるのに使っていないデータで確かめる

- 第 1 回の抗菌薬の例でも、学習に使っていない分子で確かめた。予測の上位 99 分子のうち 51 個、下位 63 分子のうち 2 個が、実際に効いた

自分の実験では、何で確かめたか

合成：反応条件の最適化

決めた　ロット 1 で温度と時間を決めた

確かめた　ロット 2（別の日・別の原料ロット）でも同じ収率が出るか

材料：組成と性質の関係

決めた　組成 5 点で傾向の式を作った

確かめた　式を作るのに使っていない組成で、性質が予測どおりか

- 「うまくいった」と書く前に、決めるのに使ったデータか、別のデータかを書いておく

自分の研究（まだなら学生実験）で、条件を決めた実験と確かめた実験が同じものを探しておく（宿題でよい）

まとめ：今日できるようになったこと

今日できるようになったこと

1. 記録してから、使っていない分子で確かめ、MAE で比べられる

   この 5 分子：M0 11.6 ℃ → M1 2.4 ℃

2. MAE は平均で上限ではなく、見て直した式（データ漏洩）や合わせすぎた式（過学習）は同じデータでよさを示せないと言える

3. 同じ記述子の異性体は区別できず、資料値のある組み合わせの外（適用範囲の外）は信用しにくいと言える

まだ分からないこと

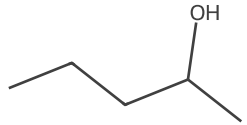
- 確かめていない組み合わせの候補を、どう扱うか → 第 6 回
- 水溶性も見て、どれを選び、次に何を測るか → 第 6 回

次回予告：第 6 回 分子候補と次の実験を提案する

第 6 話　報告する

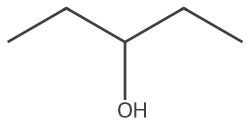


工程担当者　「で、どれを使えばいい？ 次に何を測る？」



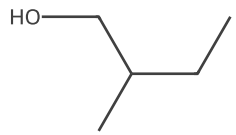
2-ペンタノール

沸点 119 ℃　w_{25} ？



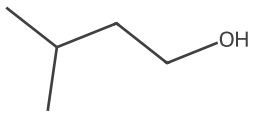
3-ペンタノール

沸点 116 ℃　w_{25} ？



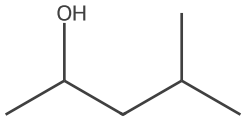
2-メチル-1-ブタノール

沸点 128 ℃　w_{25} ？



3-メチル-1-ブタノール

沸点 132 ℃　w_{25} ？



4-メチル-2-ペンタノール

$(6, 1, 1)$

沸点 ？　w_{25} —

次回の問い：複数の要求と未確認事項を踏まえ、次に何を確かめるか

今日の復習 小テスト第 5 回（manaba）・基本演習を解き直す　**教科書で読む** §5.1〜5.7

次回 第 1 回の応用演習の答えを持参　**予習** §6.1〜6.2