データから予測式を作る

分子設計 第 4 回 ｜ 教科書 第 4 章

第 4 話　式にまとめ、予測に封をする

構造と沸点の傾向を、簡単な式にまとめられるか

これまでの経過

第 3 回　「似ている」をどう定義すると、予測がどう変わるか

2-ペンタノール $(5, 1, 0)$ の最近傍予測

決め方	最近傍	予測
炭素数だけ	1-ペンタノール	**138 ℃**
$\alpha = 2$	2-ブタノール	**100 ℃**
$\alpha = 1$	2 分子が同じ距離 → 平均	**119 ℃**

$d = |n - n'| + \alpha\,|s - s'| + |b - b'|$

今日できるようになったこと

- ○ 記述子の違いから距離を計算し、最近傍の値を使って予測できる
  同じ距離の最近傍は、すべての値を平均する
- ○ 重み（何の違いを重く見るか）を変えると、最近傍も予測も変わることを示せる
- ○ 重みを少し変えただけで最近傍が入れ替わり、予測が急に変わる（138 → 100 ℃）ことを説明できる。最近傍が選べても、予測が正しいとは言えない

まだ分からないこと

- ？ 似た分子を選ぶ代わりに、構造と沸点の傾向を式にできるか → 第 4 回
- ？ 3 つの予測のどれが資料値に近いか → 第 5 回

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**　距離 $d = |n - n'| + \alpha|s - s'| + |b - b'|$、$\alpha = 1$。(4, 1, 0) の予測は？

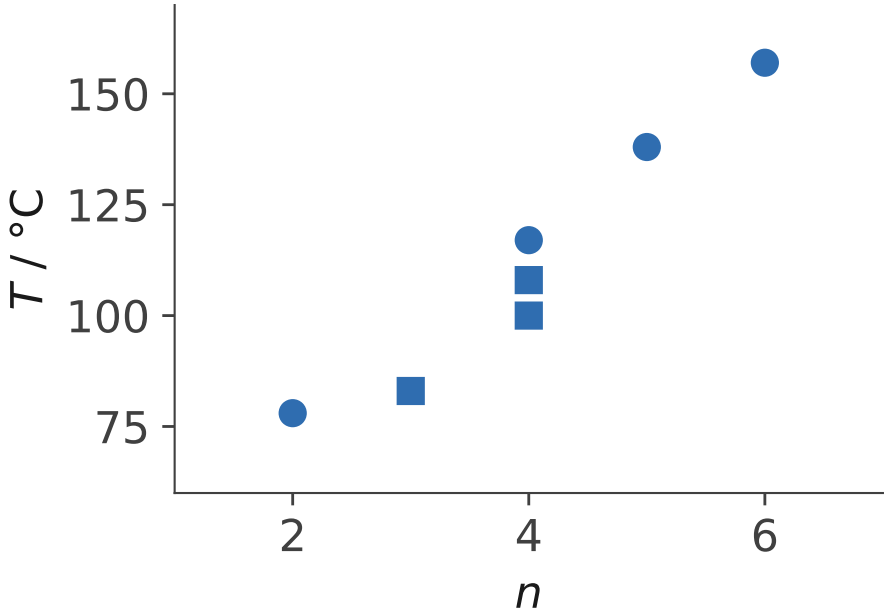
参照分子 U (3, 1, 0) 90 ℃、V (4, 0, 0) 110 ℃（どちらも架空）

(90 ℃／100 ℃／110 ℃)

問い 1　構造と沸点の傾向を、簡単な式にまとめられるか



工程担当者　「使う分子しだいで答えが変わるのは困る。式で言える？」



丸：直鎖一級。四角：二級・分岐

やってみる：丸の 4 点に直線を引くなら、
炭素 1 個あたり何 °C 上がるか。
目で読んで手元に書く

今日の流れ

教科書 §4.1〜4.6

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		

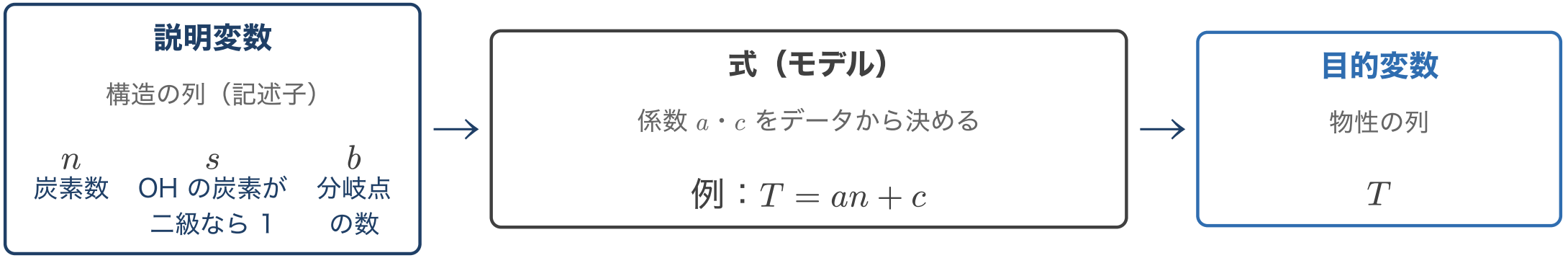
1. **問い 1 ①** 傾向を、どう式にするか → 例題・基本演習 A
2. **問い 1 ②** 式の係数を、データでどう決めるか → 例題・基本演習 B
3. **問い 1 ②** 式で予測を記録する → 応用演習 1・2

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

構造の列から、物性の列を出す式

教科書 §4.1

問い 1 ① 傾向を、どう式にするか

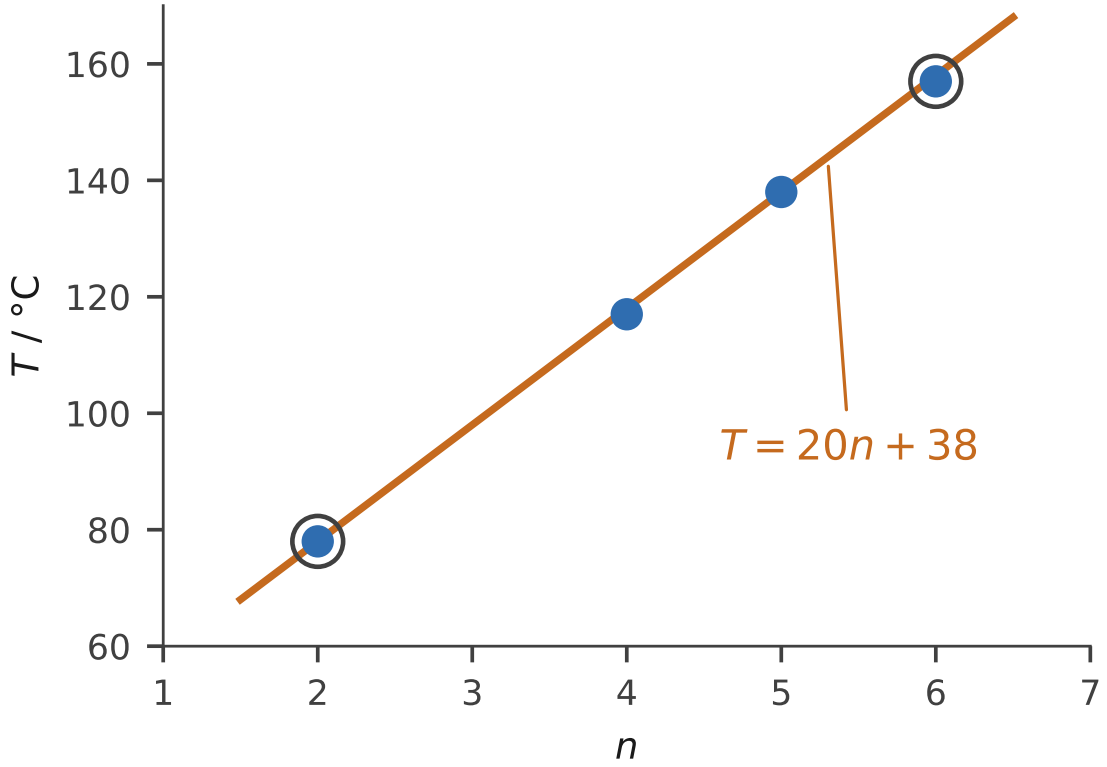


説明変数：式に入れる列　**係数**：式の中の数（a、c）。データから決める

炭素数の傾向を、直線の式にする：M0

教科書 §4.2

問い 1 ① 傾向を、どう式にするか



直鎖一級の 4 分子。丸で囲んだ両端の 2 点から傾きを求めた

- 傾き：$\dfrac{157-78}{6-2}=19.75\to 20$
- 切片：エタノールを通るように
  $78-20\times 2=38$

$$T=20n+38 \quad (\mathrm{M0})$$

M0：両端の 2 点から決めた、炭素数だけの式

例題：1-ブタノールの残差

問い 1 ①　式の予測は、どれだけずれるか

問題　1-ブタノール (4, 0, 0)、資料値 117 ℃ の M0 予測と残差を求めてください
残差＝資料値 − 予測値

解き方

① 予測　$T = 20 \times 4 + 38 = 118$ ℃

② 残差　$117 - 118 = -1$ ℃

③ 読む　負の残差：資料値が予測より 1 ℃ 低い（点が直線より下。式が高すぎる）

答え　M0 の予測 118 ℃、残差 −1 ℃

基本演習 A：M0 の残差を求める

問い 1 ①　式の予測は、どれだけずれるか

M0：$T = 20n + 38$　残差＝資料値 − 予測値

手順 1　2-プロパノール (3, 1, 0)、83 ℃ の予測と残差を手で求める

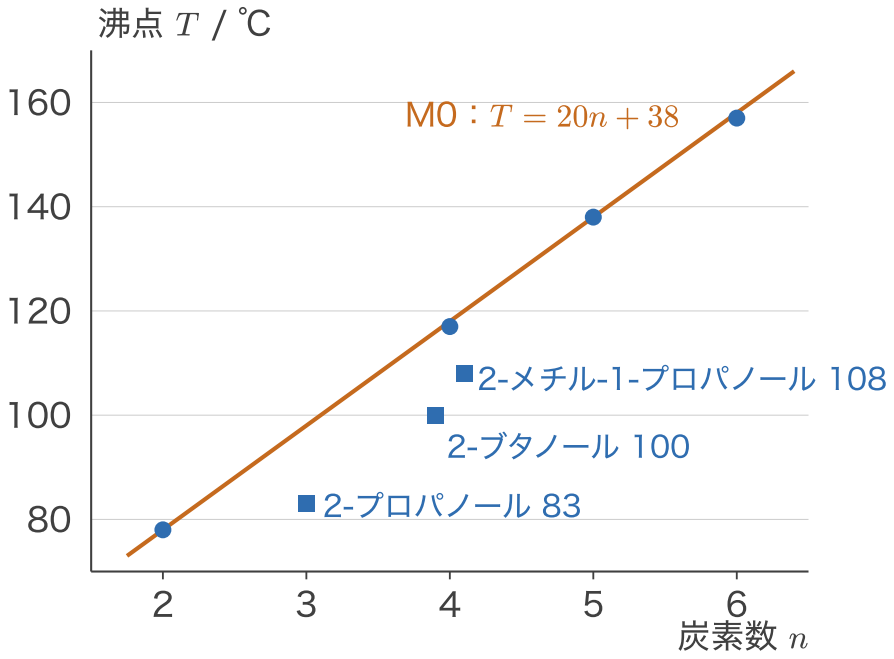
手順 2　残り 2 つの残差の符号を、図の点の位置から予想し、アプリで確かめる

2-ブタノール (4, 1, 0) 100 ℃

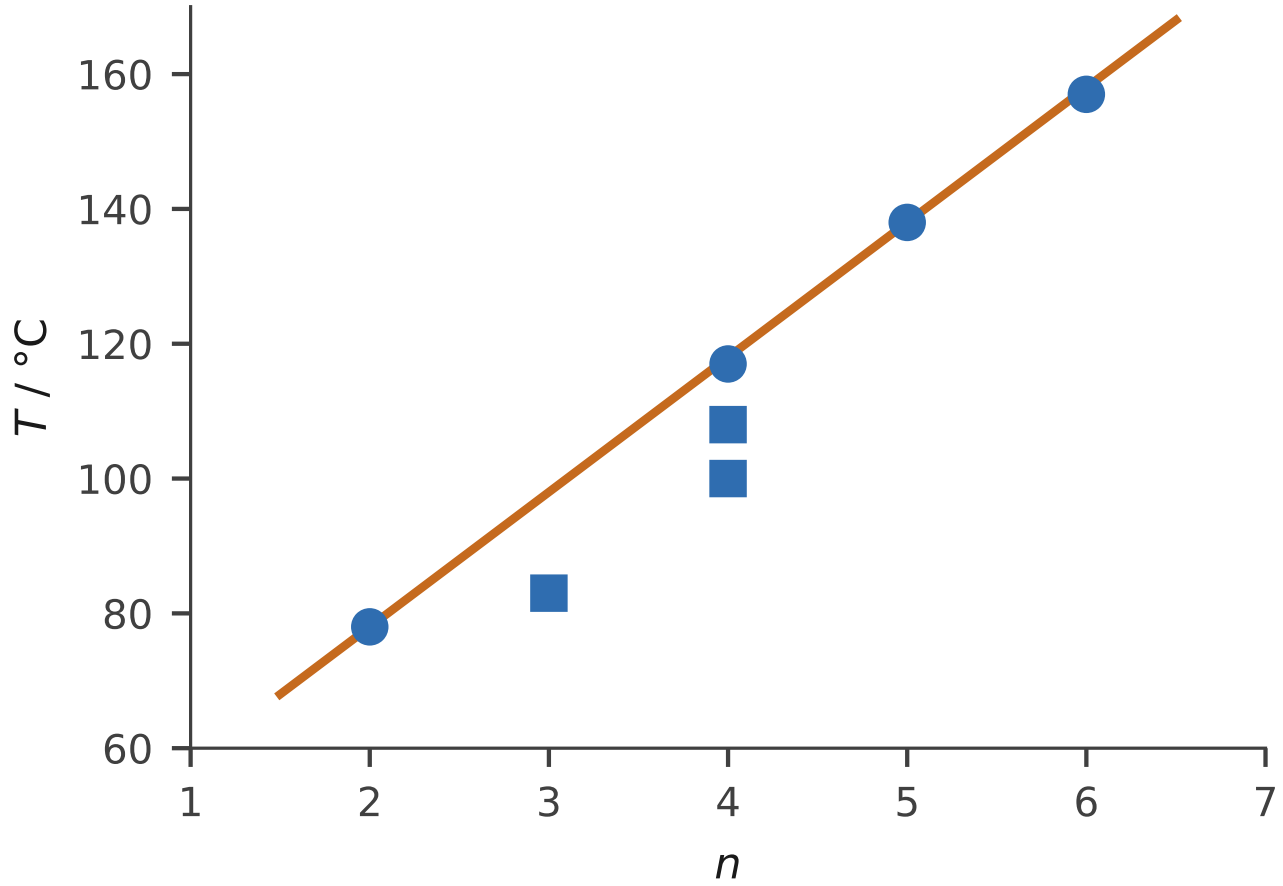
2-メチル-1-プロパノール (4, 0, 1) 108 ℃

問　3 つの残差の符号はどれも（正／負）

s・b は、直鎖一級（どちらも 0）と何が違うか



基本演習 A：残り 2 つをアプリで



四角は 2-プロパノール（n = 3、83 ℃）、2-ブタノール（n = 4、100 ℃）、2-メチル-1-プロパノール（n = 4、108 ℃）
直線 M0 との縦のずれが残差

終わったら：前の枚の問へ

基本演習 A の答え（1）

手順 1　2-プロパノールの M0 予測と残差

① 予測　$20 \times 3 + 38 = 98$ ℃

② 残差　$83 - 98 = -15$ ℃

答え　予測 98 ℃、残差 −15 ℃

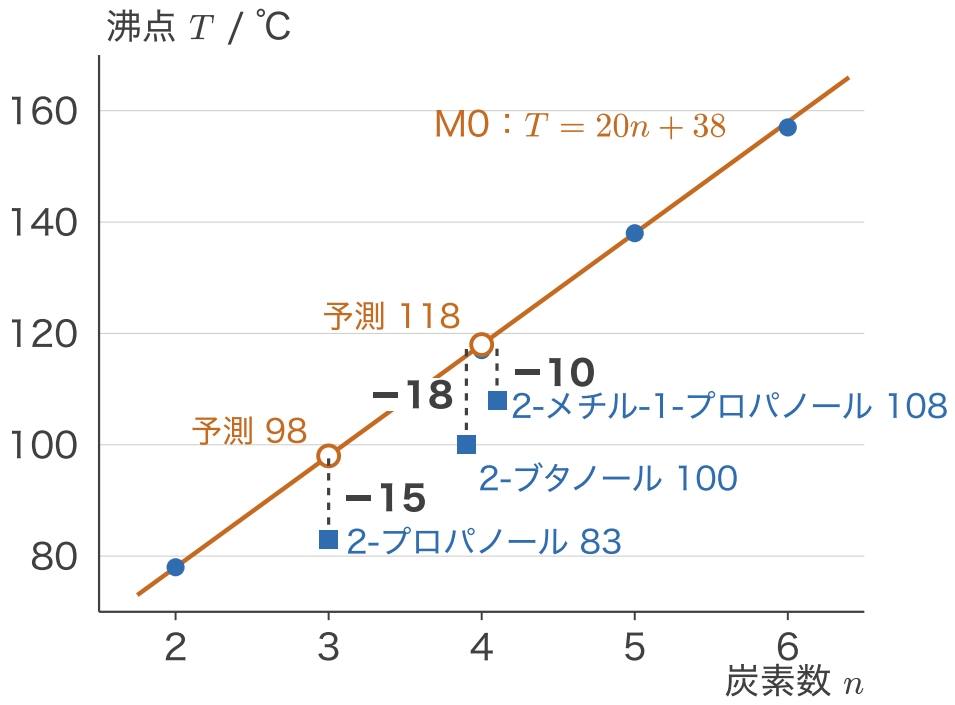
手順 2　ほかの 2 つの残差

① 2-ブタノール　$100 - 118 = -18$ ℃

② 2-メチル-1-プロパノール

$108 - 118 = -10$ ℃

答え　−18 ℃ と −10 ℃



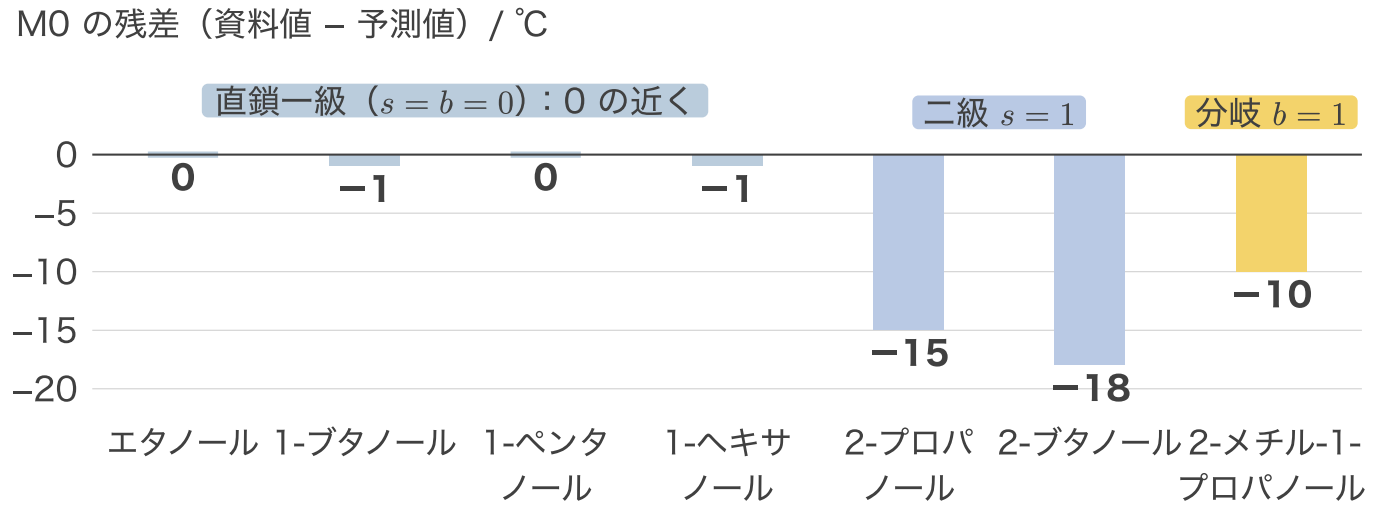
基本演習 A の答え（2）

問　3 つの残差の符号と、構造の違い

① 符号　　–15、–18、–10 ℃：3 つとも負

② 構造　　3 つとも、二級（$s = 1$）か分岐（$b = 1$）

答え　負。この二級・分岐の 3 分子を、MO は高く予測した

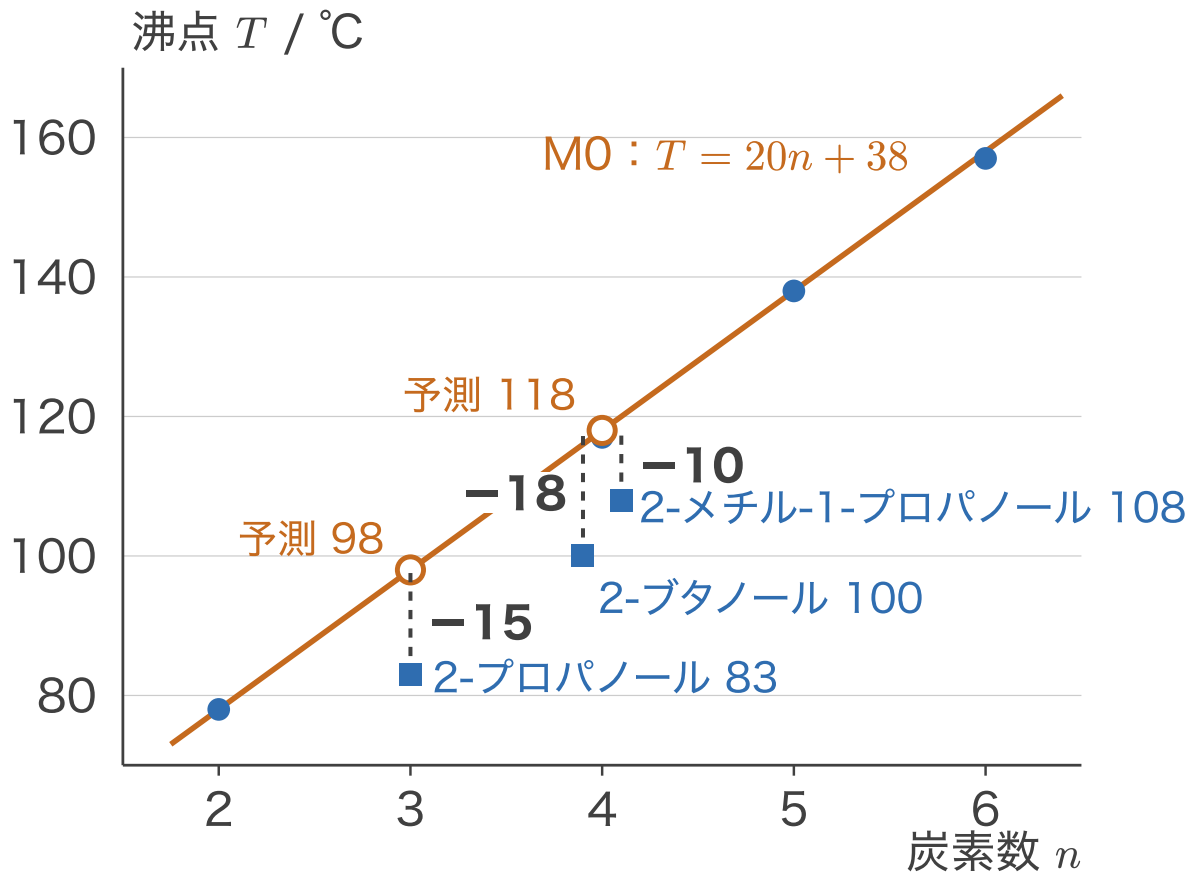


要点：残差が大きく同じ向きにそろう分子には、式が表していない構造がありそう

残差から、式を直す

教科書 §4.4

問い 1 ② 残差から、式をどう直すか



下にずれた分を、式から引く

$$T = 20n + 38 - c_s s - c_b b$$

c_s 二級の補正（℃）　c_b 分岐の補正（℃）

$s = 1$ の分子だけ c_s を引く（b も同じ）

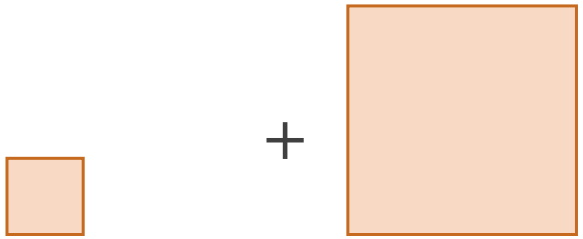
係数を、どう選ぶか：残差平方和

教科書 §4.4

問い 1 ②　係数をどう選ぶか

$$\mathrm{SSE} = e_1^2 + e_2^2 + \cdots$$

SSE 残差平方和（℃²）　$e_1, e_2, \ldots$ 各分子の残差（資料値 − 予測値）



SSE = 1 + 9 = 10

そのまま足すと +1 − 3 = −2（打ち消し合う）

例の数（どの分子の値でもない）。大きさ 3 の残差は、大きさ 1 の 9 倍に数える

係数の候補ごとに SSE を求め、SSE が最小の係数を選ぶ

例題：$c_s = 20$ の残差平方和

問い 1 ②　係数をどう選ぶか

> **問題**　二級の 2 分子（2-プロパノール (3, 1, 0) 83 ℃、2-ブタノール (4, 1, 0) 100 ℃）について、$c_s = 20$ の式 $T = 20n + 38 - 20s$ の SSE を求めてください

解き方

① 予測　2-プロパノール (3, 1, 0)：$98 - 20 = 78$
2-ブタノール (4, 1, 0)：$118 - 20 = 98$

② 残差　$83 - 78 = +5$、$100 - 98 = +2$

③ SSE　$5^2 + 2^2 = 25 + 4 = 29$ ℃²

答え　SSE = 29 ℃²（引きすぎて残差が正）

基本演習 B：SSE で c_s を選ぶ

問い 1 ②　係数をどう選ぶか

$T = 20n + 38 - c_s s$　2-プロパノール (3, 1, 0) 83 ℃、2-ブタノール (4, 1, 0) 100 ℃

手順 1　$c_s = 17$ で、2 分子の予測を出す

手順 2　残差を出す

手順 3　2 乗して足す

手順 4　最小を予想してから、アプリで 15・17・20 を確かめる　（15／17／20）

手順 5　$c_s = 16$ の SSE も手で出す（アプリに 16 はない）

問　3 つの中で最小の値だけが、最もよい係数と言えるか　（言える／言えない）

基本演習 B：アプリで確かめる

分子	(n, s, b)	資料値 / ℃	$c_s = 15$	$c_s = 17$	$c_s = 20$
2-プロパノール	$(3, 1, 0)$	**83**			78（+5）
2-ブタノール	$(4, 1, 0)$	**100**			98（+2）
SSE / ℃²					**29**

式 $T = 20n + 38 - c_s s$。$c_s = 15$・17 の列に、予測（残差）を書き込み、SSE を比べる

終わったら：前の枚の手順 5 と問へ

基本演習 B の答え（1）

手順 1〜3　$c_s = 17$ の SSE

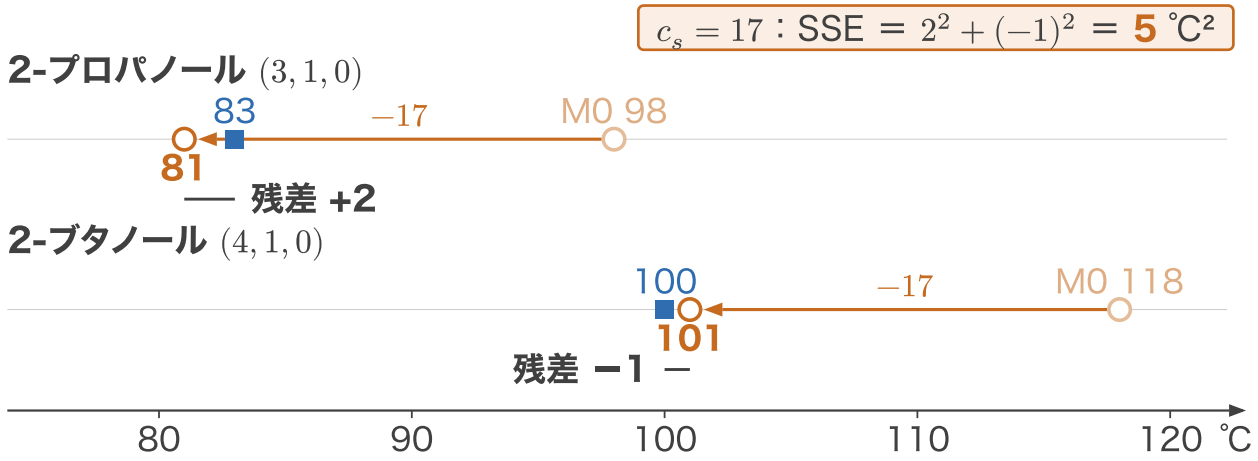
① 予測　$98 - 17 = 81$、$118 - 17 = 101$

② 残差　$83 - 81 = +2$、$100 - 101 = -1$

答え　SSE = 4 + 1 = 5 ℃²

手順 4　3 つの中で最小

答え　$c_s = 17$（SSE 5 ℃²）
15 は 9、20 は 29 ℃²



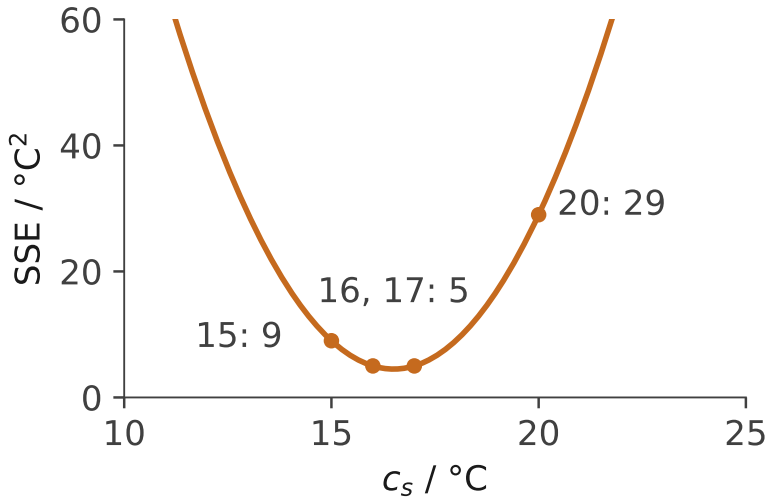
基本演習 B の答え（2）

手順 5・問　17 だけが最もよい係数と言えるか

① 16　予測 82・102、残差 +1・−2、SSE ＝ 1 ＋ 4 ＝ 5 ℃²

② 最小　連続の値で SSE が最小なのは 16 と 17 の間（16.5。教科書 §4.6）

答え　言えない。16 でも SSE は同じ 5 ℃²（17 は試した 3 つで最小なだけ）

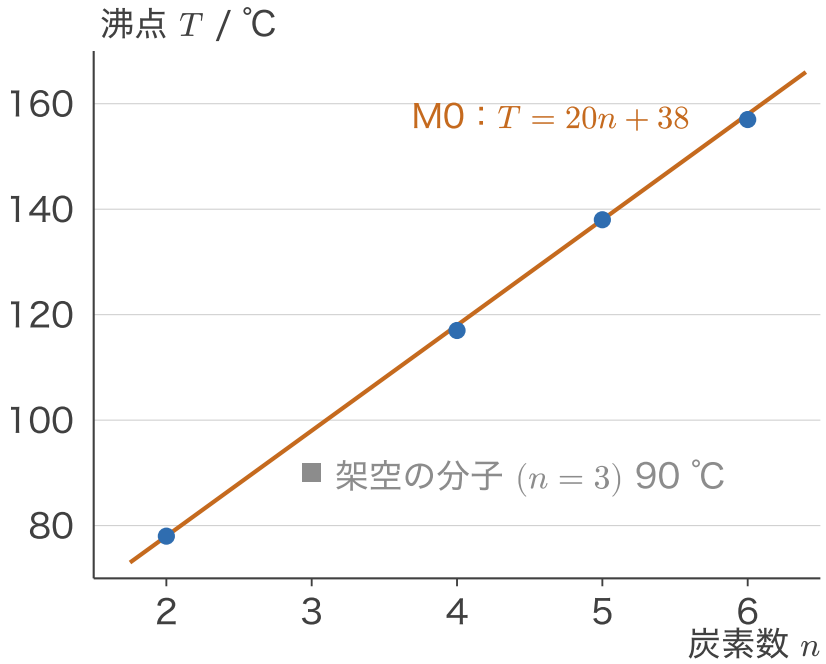


要点：試した 3 つの中で SSE が最小の 17 を、以後使う。選び方を書いておく

理解確認（個人・3 分）

M0：$T = 20n + 38$。残差＝資料値 − 予測値

1. 炭素数 3 の架空の分子の沸点が 90 ℃。M0 の予測と残差を求める
2. 二級の補正の係数 c_s は 17。「二級にすれば、どの分子でも必ず沸点が 17 ℃ 下がる」と言えるか
   （言える／言えない）
   理由を 1 文で



理解確認の答え

問 1　炭素数 3・90 ℃ の分子の M0 予測と残差

① 予測　$20 \times 3 + 38 = 98$ ℃

② 残差　$90 - 98 = -8$ ℃（点が直線より下）

よくある誤り　$98 - 90 = +8$：引く向きが逆（残差＝資料値 − 予測値）

答え　予測 98 ℃、残差 −8 ℃

問 2　「二級にすれば、どの分子でも必ず 17 ℃ 下がる」と言えるか

考え方　17 は 2 分子に合わせて選んだ値。その 2 分子でも残差は −15 と −18 で違い、ほかの分子は確かめていない

よくある誤り　「データから出した値だから、一般に成り立つ」：係数を法則として読んでいる

答え　言えない

応用演習 1：M1 を作る

問い 1 ②　分岐の係数をどう決めるか

2-メチル-1-プロパノール (4, 0, 1)：M0 の予測 118 ℃、資料値 108 ℃（残差 −10 ℃）

手順 1　残差が 0 になる c_b を求める：$118 - c_b = 108$

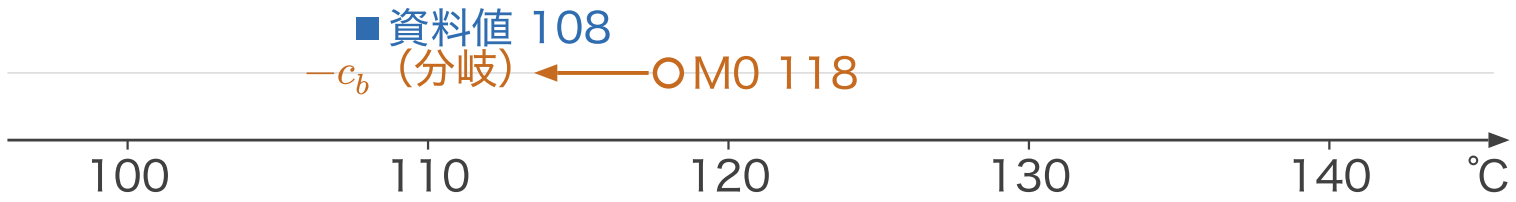
手順 2　$c_s = 17$ とあわせて式を書く（この式を **M1** と呼ぶ）：

$$T = 20n + 38 - 17s - c_b b$$

分岐の分子は 1 つだけ。1 分子なら、SSE が最小になるのは残差が 0 のとき

M1 = M0 − 17 s − c_b b

2-メチル-1-プロパノール
$(4, 0, 1)$



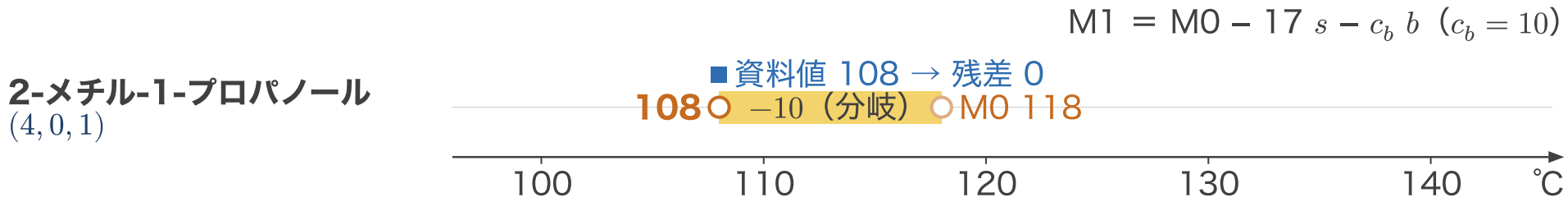
応用演習 1 の答え

M1 を作る

① c_b　　$118 - c_b = 108 \rightarrow c_b = 10$（1 分子なので、残差 0 で SSE が最小）

② **M1**　　$T = 20n + 38 - 17s - 10b$

答え　$c_b = 10$



応用演習 2：予測を記録する

問い 1 ②　予測をどう記録するか

M1：$T = 20n + 38 - 17s - 10b$

分子	最近傍 炭素数だけ	最近傍 $\alpha = 2$	最近傍 $\alpha = 1$	M1
2-ペンタノール (5, 1, 0)	138 ℃	100 ℃	119 ℃	
3-メチル-1-ブタノール (5, 0, 1)	—	—	—	

—：第 3 回で求めていない

手順 1　表を手元に写し、M1 で 2 つを予測して書く

手順 2　封をする（資料値を見るまで書き換えない）

問　「補正の係数を決めた 3 分子（2-プロパノール、2-ブタノール、2-メチル-1-プロパノール）で残差が小さい。だから、ほかの分子でもよく当たる」と言えるか（言える／言えない）

応用演習 2 の答え（1）

M1 の予測

① **2-ペンタノール** $20 \times 5 + 38 - 17 = 121$ ℃

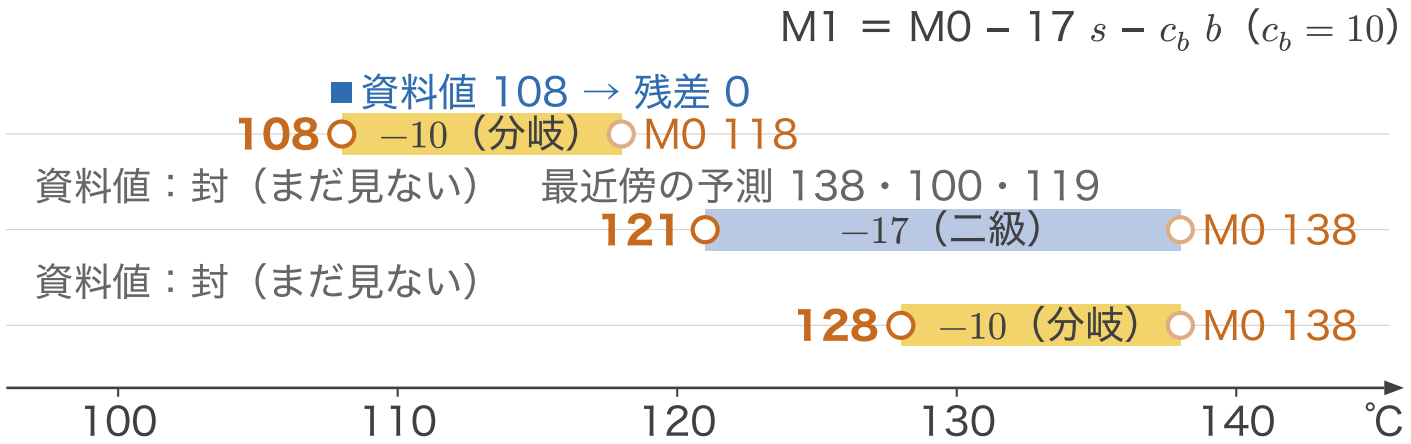
② **3-メチル-1-ブタノール** $20 \times 5 + 38 - 10 = 128$ ℃

答え　121 ℃ と 128 ℃。表の M1 の列に書いて封をする

2-メチル-1-プロパノール
$(4, 0, 1)$

2-ペンタノール
$(5, 1, 0)$

3-メチル-1-ブタノール
$(5, 0, 1)$



応用演習 2 の答え（2）

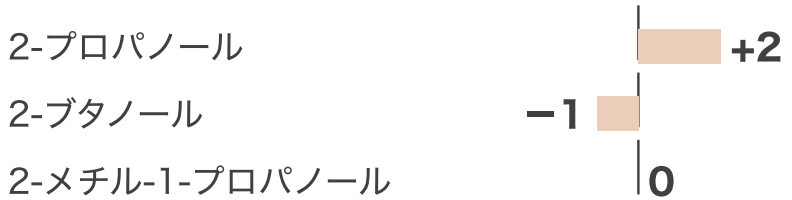
問　補正の係数を決めた分子で合えば、ほかの分子でもよく当たるか

考え方　残差（+2、−1、0）が小さいのは、その 3 分子に合わせて係数を選んだから

成立しない例　「残差が小さいので、M1 は正しい式だ」：使っていない分子で確かめていない

答え　言えない。係数を決めるのに使っていない分子で確かめる（第 5 回）

補正の係数を決めた 3 分子（当てはまり）
M1 の残差（資料値 − 予測値）/ ℃



この 3 分子に合わせて係数を選んだ → 小さくて当然

使っていない分子（予測）
M1 の予測と残差

2-ペンタノール　予測 **121** ℃　残差 ?

3-メチル-1-ブタノール　予測 **128** ℃　残差 ?

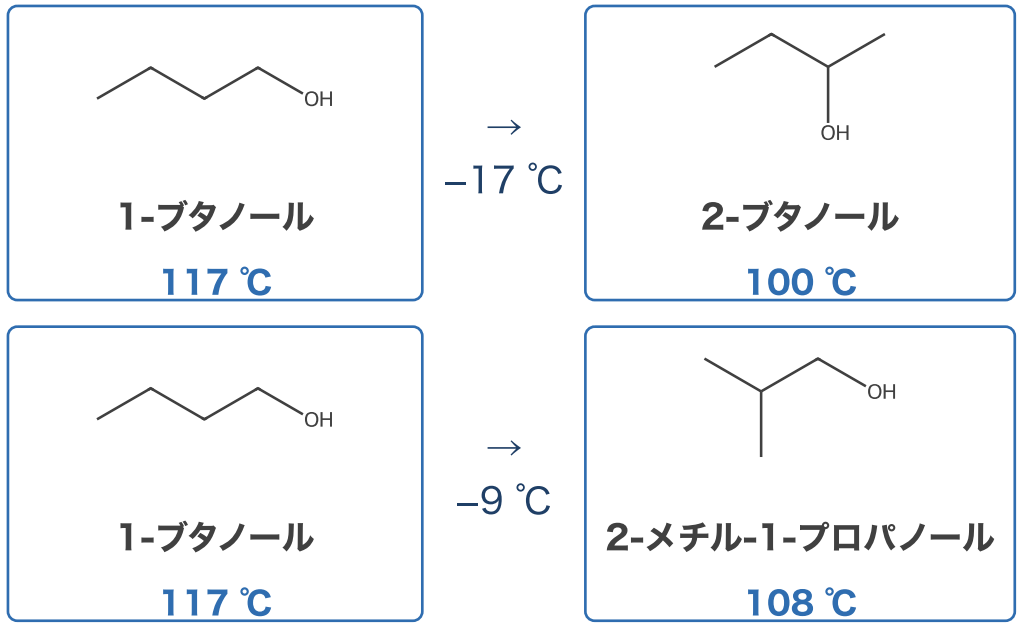
資料値は封をしたまま → 第 5 回で確かめる

要点：式がデータに合うことと、未知の分子で当たることは別。係数は法則ではない

係数の向きを、化学で説明する

教科書 §4.6

問い 1 ② 係数の向きを、どう説明するか



二級：M1 では $-17s$（$c_s = 17$）

OH の炭素の周りの立体障害 → 水素結合をつくりにくい

分岐：M1 では $-10b$（$c_b = 10$）

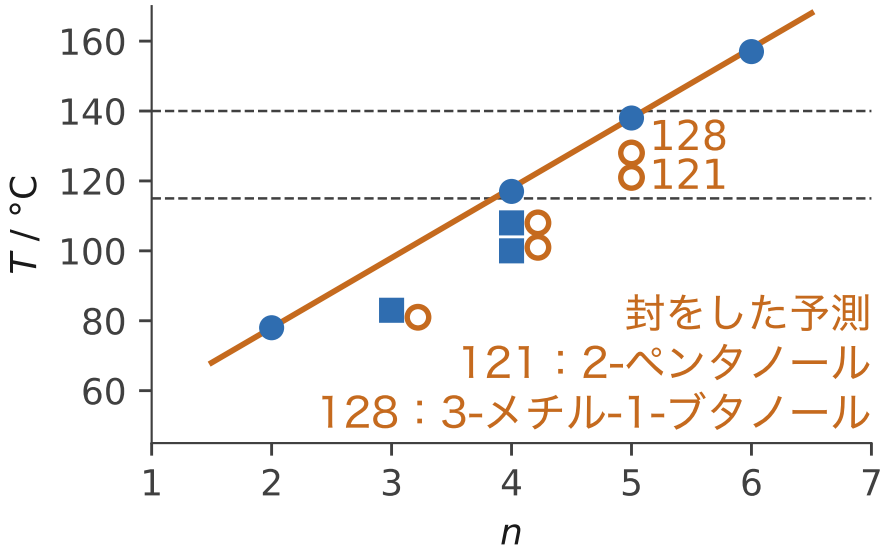
接する面が減る → 分散力が弱い

−17・−9 ℃ は 1-ブタノールとの差（資料値）。係数 17・10 は M0 の残差から決めた別の数

下がる向きは分子間の力で説明できる（仮説）。どれだけ下がるかは決まらない

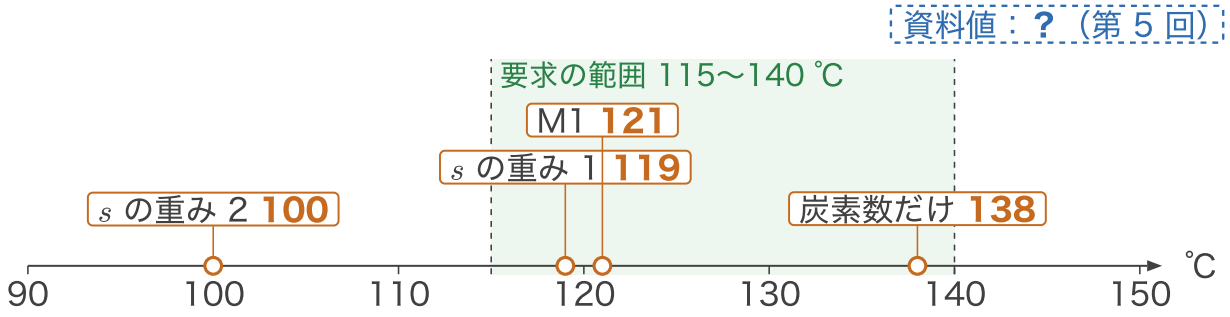
問い 1 の答え　7 分子の傾向は、炭素数の直線＋補正（M1）で表せる

問い 1：構造と沸点の傾向を、簡単な式にまとめられるか



四角：資料値　白丸：M1 の予測　破線：範囲

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



2-ペンタノールの 4 つの予測

- 傾きと切片は端の 2 点、c_s は 2 分子の SSE、c_b は 1 分子の残差から決めた。当たるかは第 5 回に確かめる

実際のケモインフォでは：データから式を決める

この授業（今日の M1）：沸点

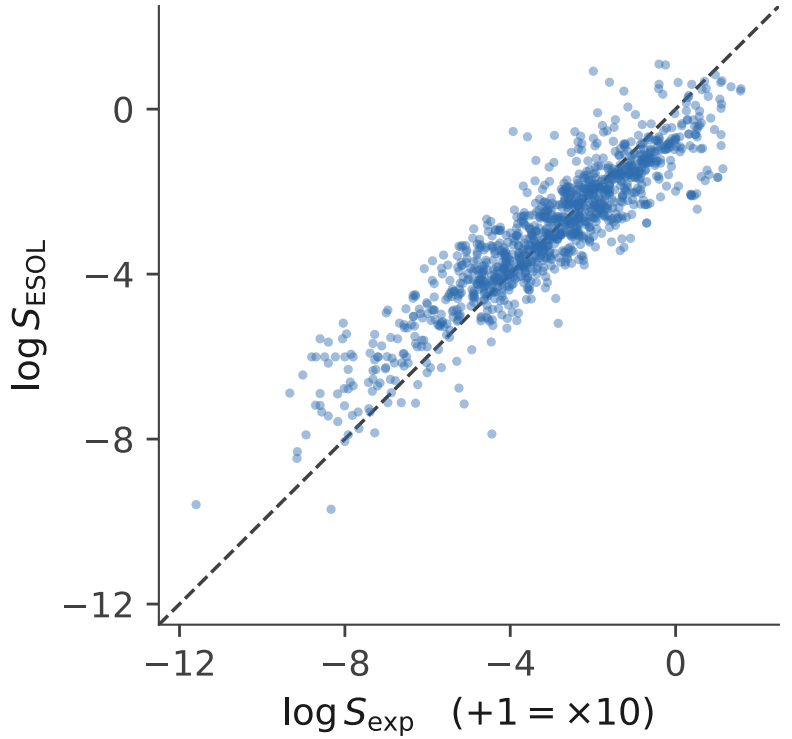
$$T = 20n + 38 - 17s - 10b$$

記述子　3 つ（炭素数・OH の炭素の級・分岐点の数）
決め方　直鎖の 2 点と二級・分岐の 3 分子から、手で決めた

実際の研究：水溶性の予測式 ESOL（Delaney 2004）

$$\log S = 0.16 - 0.63\,\mathrm{clogP} - 0.0062\,\mathrm{MW} + 0.066\,\mathrm{RB} - 0.74\,\mathrm{AP}$$

記述子　4 つ（油と水への分かれやすさ・分子量など）
S　水への溶解度（mol/L）
決め方　測定値から、回帰（SSE が最小になる係数をプログラムで求める）で決めた



ESOL の式の値と測定値（1,128 分子。S は mol/L）
破線の上なら、式の値＝測定値

M1 と同じく、式の係数をデータから決める（機械学習でいう「学習」の簡単な形）

自分の実験では、式から何が言えるか

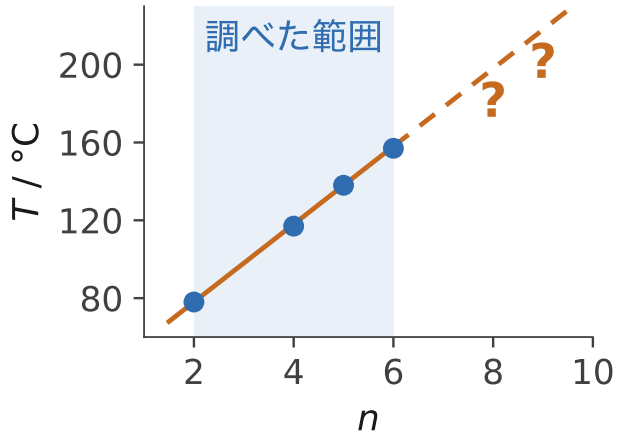
例（架空）：焼成温度を変えた試料で、温度が高いほど電気伝導度が高い傾向。直線の式に当てはめた

式から言えること

・調べた温度の範囲での、平均的な傾向
・その範囲の中の、別の温度での見積もり

式だけでは言えないこと

・なぜ上がるのか（機構）
・範囲の外の温度でも同じ傾きか
・温度を上げれば、必ず上がるか



今日の MO を、調べた範囲の外へ延ばした

- 自分の研究（まだなら学生実験）で、傾向を式や直線で表しているものを 1 つ挙げ、言えないことを考えておく（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. 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 ℃

2. 残差がそろう分子から補正の項を入れ、係数を SSE で選べる

   c_s は 15・17・20 から 17（16 も同じ SSE）

3. 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 ℃

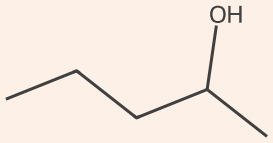
次回予告：第 5 回 予測を確かめ、信用できる範囲を考える

教科書 第 5 章

第 5 話　封を開ける



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？」



2-ペンタノール

資料値 ？

予測の方法	回	予測
最近傍（炭素数だけ）	第 3 回	**138 ℃**
最近傍（$\alpha = 2$）	第 3 回	**100 ℃**
最近傍（$\alpha = 1$）	第 3 回	**119 ℃**
M1	第 4 回	**121 ℃**

次回の問い：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

持ち物：今日の応用演習で封をした予測の記録（表）

復習 小テスト第 4 回（manaba）・基本演習を解き直す　**教科書で読む** §4.1〜4.6・付録 A　**予習** §5.1〜5.2