第96回数学教育実践研究会　レポート発表

確率・統計でOne more thing

北海道札幌南高等学校教諭　長尾良平

平成28年1月30日　ニッセイMKビル

1　はじめに

本校では, 1年生で数学I, 数学II, 数学Aを合計6単位で実施している．受験を意識するため進度は速いが, 生徒の印象に残るような授業を行いたいと日々考えている．

Apple社の故Steve Jobs氏は製品発表のプレゼンの終盤に **“One more thing”** と呟き, 目玉製品を紹介するのがお約束だった．

本発表では, 教科書の内容 $+\alpha$ としての “One more thing” を確率・統計の単元での実践例から紹介していきたい．

2　同様に確からしい

高校では, **「同様に確からしい」** という概念に基づいて確率が定義される．その意味を確認するために, 有名な逸話を取り上げた．

ヨーロッパの貴族の話

サイコロを3個同時に投げて, 出た目の和を考える．
☆和が9になるのは,
⚀⚁⚅, ⚀⚂⚄, ⚀⚃⚃, ⚁⚁⚄,
⚁⚂⚃, ⚂⚂⚂の6通り．
☆和が10になるのは,
⚀⚂⚅, ⚀⚃⚄, ⚁⚁⚅, ⚁⚂⚄,
⚁⚃⚃, ⚂⚂⚃の6通り．

どちらに賭けても平等な筈だが, 体感的には後者に賭けた方が有利な気がする・・・．

何人かの生徒にどう考えれば良いか聞いていくうちに, 「⚀⚁⚅は6通り, ⚀⚃⚃は3通りの可能性があり, **起こりやすさが異なる**」との意見が出てくる．

数え上げてみると和が9の場合が25通り, 10の場合が27通りあり, **「どちらに賭けても平等ということが間違い」** ということが分かる．それぞれの確率が大体0.116と0.125であり, 1％しかない差を感じていたことに呆れつつこの話を終える．また, **一昨年の夏休みに実家で2000回サイコロを投げた結果**も併せて紹介する．以前のレポート[1]同様, 生徒には呆れられる．

3　反復試行

ひととおり反復試行の演習が終わった段階で, 2つの問を扱った．元ネタは[2]である．

効果測定について

自動車教習所では, 検定を受けるための要件の1つとして **「効果測定」** という筆記試験があります．効果測定は**○×の2者択一式50問**の試験で, **45問以上正解で合格**となります．

不勉強なA君が**鉛筆を転がしてテキトーに答えて**いった（つまり, ○×を確率 $\frac{1}{2}$ で選ぶ）ときに, 合格できる確率は**およそ**どれくらいか？

（注）$2^{10} = 1024 \fallingdotseq 1000$ を用いてもよい．

$$\sum_{k=45}^{50} {}_{50}\mathrm{C}_k \left(\frac{1}{2}\right)^{50} \fallingdotseq \frac{2.3\times 10^6}{10^{15}} \fallingdotseq \frac{1}{500000000}$$

となり, **「まじめに勉強した方がいい」**という**至極真っ当な結論**が得られる. 続けて, 次の問である.

アラビア語のテスト

> 次のアラビア語の意味を考え, 正しいと思う方に〇をつけなさい（問題は別紙参照）.
> 正解が 14 問以下又は 36 問以上の人にはチョコレート謹呈.

問題用紙は, Google 翻訳の結果を Excel に貼り付けて作成した. 生徒は見たことがないアラビア語の羅列に面食らいつつも, **効果測定の問題よりも合格ゾーンが広い**こともあって何とか景品にありつこうと必死になって取り組んでいた.

テスト終了後に採点を行い, 翌日の授業で返却した. クラス毎の素点分布の入った講評も配り, **「何でもっと悪い点数が取れないんだ!!」**とひとしきり説教をし, $\mathrm{P}(X \leqq 14, 36 \leqq X) \fallingdotseq 0.003$ であることや,「千三つ」という言葉について紹介した.

昨年も同じテストを行ったが, 景品をゲットできた生徒はいなかった. また, 昨年の宿泊研修のバスの中で**「塾で藻岩の子に『南ってアラビア語のテスト有るんだよね〜』って聞かれたんだぁ」**という生徒の会話が聞こえてきたときには, 思わず吹き出してしまった.

講評の得点分布を見ると, 25 問付近に正答数のピークがくる. 問題数が 50 で当たる確率が半々であるから, 尤もらしい結論ではあるが, 次の問で確認しておいた.

正答数のピークは？

> 問　50 問の 2 択のクイズにテキトーに答えるとき, k 問正解したときの確率を p_k で表す. p_k が最大となる k の値は？

$p_k = {}_{50}\mathrm{C}_k \left(\frac{1}{2}\right)^{50}$, $p_{k+1} = {}_{50}\mathrm{C}_{k+1} \left(\frac{1}{2}\right)^{50}$ より

$$\frac{p_{k+1}}{p_k} = \frac{{}_{50}\mathrm{C}_{k+1}\left(\frac{1}{2}\right)^{50}}{{}_{50}\mathrm{C}_k\left(\frac{1}{2}\right)^{50}} = \frac{{}_{50}\mathrm{C}_{k+1}}{{}_{50}\mathrm{C}_k}$$
$$= \frac{50!}{(k+1)!(49-k)!} \times \frac{k!(50-k)!}{50!}$$
$$= \frac{50-k}{k+1}$$

(i) $\frac{p_{k+1}}{p_k} > 1$ のとき

$\frac{50-k}{k+1} > 1$ を解いて, $k < \frac{49}{2}$

k は 0 以上の整数だから,

$$p_0 < p_1 < p_2 < \ldots < p_{24} < p_{25}$$

(ii) $\frac{p_{k+1}}{p_k} < 1$ のとき

$\frac{50-k}{k+1} < 1$ を解いて, $k > \frac{49}{2}$

k は 50 以下の整数だから,

$$p_{25} > p_{26} > p_{27} > \ldots > p_{49} > p_{50}$$

(i)(ii) より, $k = 25$ のとき, p_k が最大となる.

4　条件付き確率

今回の改訂で数学 C から数学 A に移った条件付き確率も, 色々遊べる面白い単元である. まず, 有名な**モンティ・ホール問題**を扱った例から.

モンティ・ホール問題

> 区別のつかない 3 つの箱があり, その中に 1 個だけ, 当たりくじの入っている箱がある.
> まず, 解答者に箱を選んでもらう. 司会者は, どの箱に当たりくじが入っているかを知っているので, **残った 2 つの箱から当たりくじの入っていない箱**を 1 個だけ取り除く.
> 司会者は**「選んだ箱を変えるチャンスを 1 度だけあげます！ どうしますか？」**と解答者にたたみかける.
> 果たして, 解答者は箱を変えた方がいいのであろうか？ それとも, そのままの方がいいのであろうか？

座席が隣の生徒をペアにし, 紙コップを 3 個ず

つ配り, 消しゴムを当たりくじと想定させて実際にやらせてみた. 司会者と解答者の両方を体験させたが, 大盛り上がりであった.

3クラス分の正答数の集約

	変　更	維　持	合　計
総　計	709	415	1124
比　率	0.63	0.37	1

1回毎に記録を取り, 授業の終わりに提出させた. 次の時間に, 集約した結果を配布し, **「何故, 変更した方が変更しないより約2倍当たりやすいのか？」**を隣同士で考えさせた. そして, 何人かの生徒を指名して黒板等を使いながら説明してもらった. 上手く説明できなかった生徒も, 他の生徒の説明を聞いて, 納得していた.

次に, **ベイズ更新**を扱った例を紹介したい. 元ネタは [3] である.

真面目さ度合いを数値化する

過去の分析から, 次のことが分かっている.

待ち合わせに対する感覚（確率）

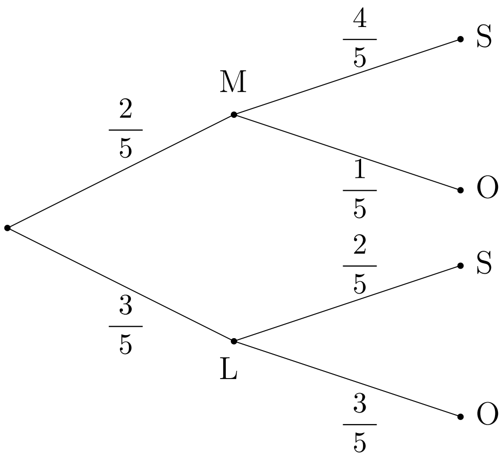
	セーフ S	アウト O
真面目な人 M	$\frac{4}{5}$	$\frac{1}{5}$
ルーズな人 L	$\frac{2}{5}$	$\frac{3}{5}$

少しルーズだと思われている A 君は, 5回のデートの待ち合わせに SSOOS だった. A 君が真面目な人と思われるかどうかを確率で考えよ.

A 君は少しルーズ：$P(M) = \frac{2}{5}$ と設定してみる.

(1) 1回目：S

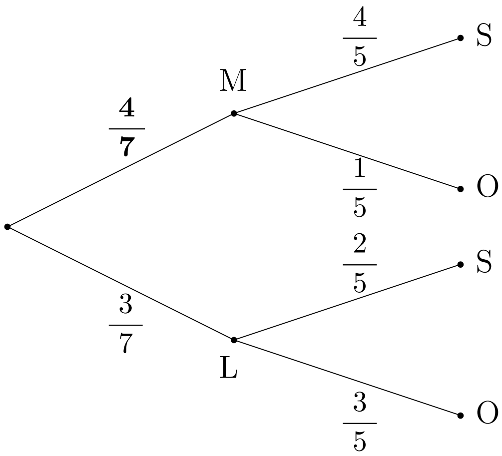
$$P_S(M) = \frac{\frac{2}{5} \times \frac{4}{5}}{\frac{2}{5} \times \frac{4}{5} + \frac{3}{5} \times \frac{2}{5}} = \frac{4}{7}$$



P(M) が $\frac{2}{5}$ から $\boldsymbol{\frac{4}{7}}$ に**更新される**.

(2) 2回目：S

$$P_S(M) = \frac{\frac{4}{7} \times \frac{4}{5}}{\frac{4}{7} \times \frac{4}{5} + \frac{3}{7} \times \frac{2}{5}} = \frac{8}{11}$$



(3) 3回目：O

$$P_O(M) = \frac{\frac{8}{11} \times \frac{1}{5}}{\frac{8}{11} \times \frac{1}{5} + \frac{3}{11} \times \frac{3}{5}} = \frac{8}{17}$$

(4) 4回目：O

$$P_O(M) = \frac{\frac{8}{17} \times \frac{1}{5}}{\frac{8}{17} \times \frac{1}{5} + \frac{9}{17} \times \frac{3}{5}} = \frac{8}{35}$$

(5) 5回目：S

$$\mathrm{Ps(M)} = \frac{\frac{8}{35} \times \frac{4}{5}}{\frac{8}{35} \times \frac{4}{5} + \frac{27}{35} \times \frac{2}{5}} = \frac{16}{43}$$

5回のデートの待ち合わせの記録から, A君の真面目さ度合いは,
40 % $\xrightarrow{S}$57 % $\xrightarrow{S}$72 % $\xrightarrow{O}$47 % $\xrightarrow{O}$22 % $\xrightarrow{S}$37 %
と変動しており, **データを取り込んで確率を更新することによって，人間の感情や感覚を的確に表現している**と言える. 生徒も, 行動の結果がその後の確率に反映されることに感心していた.

5　相関係数に絡めて

統計学は, データからいろいろな情報を導き, 推測していくところに面白みがあると感じている. 数学Iの「データの分析」は主に記述統計を扱っており, 推測統計は数学Bの内容になっている. そこで, 相関係数の応用として**最小二乗法**を扱ってみた. 難しい計算等は省略し, 「中学校で実験のデータに対して直線引いてみたよね？あれの一般化だよ」というかたちで紹介した.

桜の開花日を予想する

気象庁のWebサイトでは, 毎月の平均気温や毎年の桜の開花日などのデータがダウンロードできます.

そこで, 1971年〜2000年の「3月の平均気温」と「桜の開花日」のデータ30年分を基にして回帰直線を求め、2001年以降の開花日を予想してみました.

授業では数学者をよく取り上げるが, ガウスが何回か登場している. 最小二乗法もガウスの発明であり, **「またガウスか〜」**との声が聞こえた.

最小二乗法では

$$d = \sum_{k=1}^{n} \{y_k - (ax_k + b)\}^2$$

を最小とする a, b の値を求め, $y = ax + b$ が回帰直線の方程式となる. そのためには, 偏微分を用いるのが一番簡単だが, まだ微分を扱っていない. そこで, 平方完成でごりごりやる例（簡単にデータが2つの場合）をプリントの裏面に印刷しておいた.

$$a = \frac{s_{xy}}{{s_x}^2} = r \cdot \frac{s_y}{s_x}\ ,\ b = \overline{y} - a\,\overline{x}$$

力技だが, 一部の生徒は熱心にプリントを眺め計算を確認していた.

6　終わりに

本校の生徒は受験に向けて意識が高いが, そんな彼らにこそ, ちょっと寄り道の内容を知っておいて欲しいと思う. **確率・統計の単元は身の周りの事象との結びつきが強い**ので, いろいろと遊ぶことができる. そういった**遊び心を大切にしながら，数学について面白いなあと思って欲しい**と考えている.

また, 前回の数実研の大谷先生の発表で**「授業でこんなことやったよね・・・ああそうそう」という記憶が大事**ではないかという指摘があったが, そのとおりだと思う.

これからも, 様々な単元での “one more thing” を追求していきたい.

参考文献等

[1] 長尾良平「無駄な？ 努力」
第83回数学教育実践研究会レポート

[2] 黒田孝郎他「高等学校の確率・統計」
ちくま学芸文庫

[3] 涌井良幸他「Excelでスッキリわかるベイズ統計入門」日本実業出版社

[4] 私的数学塾
`http://www004.upp.so-net.ne.jp/s_honma/index.htm`

[5] 気象庁
`http://www.jma.go.jp/jma/index.html`

アラビア語のテスト

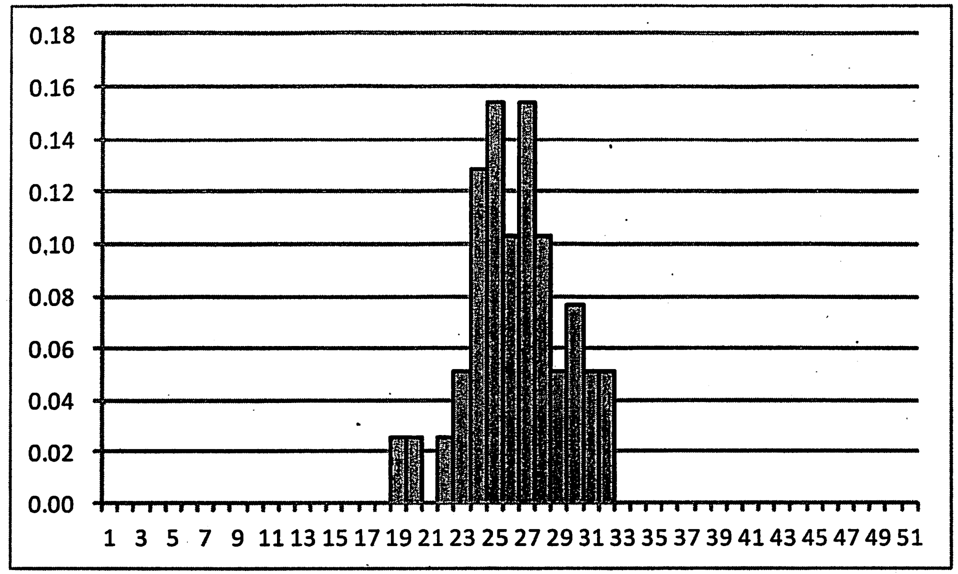
組　　　　氏名

問　次のアラビア語の意味を考え、正しいと思う方に○をつけなさい。

1	المدرسة	ア 学校	イ 家	26	الغابات	ア 林	イ 森
2	اللغة الوطنية	ア 国語	イ 数学	27	حذاء	ア 手袋	イ 靴
3	الحب	ア 愛	イ 平和	28	النار	ア 地震	イ 火事
4	البحر	ア 海	イ 川	29	وظيفة	ア 確率	イ 関数
5	الشمس	ア 太陽	イ 地球	30	غادر	ア 右	イ 左
6	المرض	ア 健康	イ 病気	31	الجبل	ア 山	イ 谷
7	سفينة	ア 車	イ 船	32	الأزرق	ア 青	イ 赤
8	الشهر	ア 星	イ 月	33	غائم جزئيا	ア 晴れ	イ 雨
9	رجل	ア 女	イ 男	34	تحت	ア 下	イ 上
10	مكتب	ア 椅子	イ 机	35	الضحك	ア 笑う	イ 泣く
11	الشيطان	ア 悪魔	イ 神	36	من الصعب	ア 臭い	イ 硬い
12	الماشية	ア 牛	イ 豚	37	القط	ア 犬	イ 猫
13	الليل	ア 夜	イ 昼	38	الشاي الأسود	ア 珈琲	イ 紅茶
14	عالية	ア 高い	イ 低い	39	خارج	ア 内	イ 外
15	على مدار الساعة	ア 時計	イ 鏡	40	الماضي	ア 未来	イ 過去
16	الحديد	ア 木	イ 鉄	41	الجامعة	ア 大学	イ 高校
17	رقيقة	ア 太い	イ 細い	42	دائرة	ア 円	イ 直線
18	كتاب	ア テレビ	イ 本	43	الأخدود	ア 溝	イ 穴
19	زهرة	ア 種	イ 花	44	اللحوم	ア 肉	イ 魚
20	العنب	ア リンゴ	イ ぶどう	45	بوذا	ア 仏	イ 神
21	الأرز	ア 米	イ 麦	46	الفضة	ア 金	イ 銀
22	ثقيلة	ア 重い	イ 軽い	47	الأذن	ア 目	イ 耳
23	صوت	ア 音	イ 光	48	ضفدع	ア 蛇	イ 蛙
24	الصخور	ア 岩	イ 砂	49	السقف	ア 床	イ 天井
25	الخمور	ア 酒	イ 水	50	نجاح	ア 失敗	イ 成功

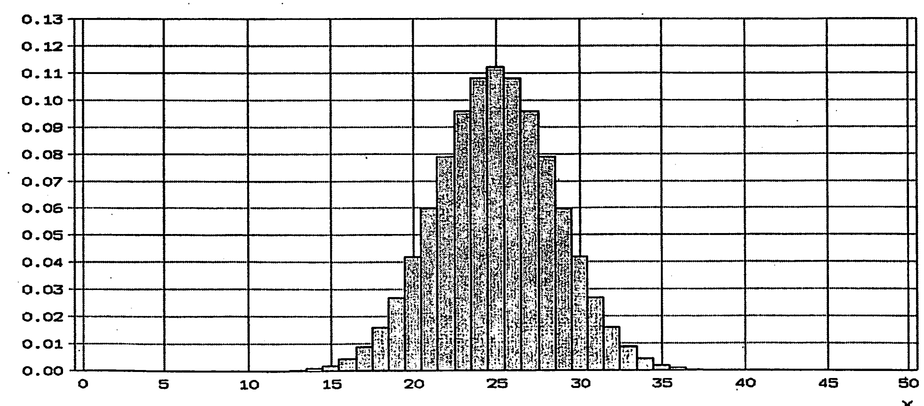
1年2組　アラビア語のテストを終えて

1．得点分布



講評：上位者、下位者ともに少ないのでは？頑張りが足りない？

2．二項分布B（50，1/2）のヒストグラム



モンティ・ホール問題　記録用紙

回数	変更した・しない		結果	
1	変更	維持	当たり	ハズレ
2	変更	維持	当たり	ハズレ
3	変更	維持	当たり	ハズレ
4	変更	維持	当たり	ハズレ
5	変更	維持	当たり	ハズレ
6	変更	維持	当たり	ハズレ
7	変更	維持	当たり	ハズレ
8	変更	維持	当たり	ハズレ
9	変更	維持	当たり	ハズレ
10	変更	維持	当たり	ハズレ
11	変更	維持	当たり	ハズレ
12	変更	維持	当たり	ハズレ
13	変更	維持	当たり	ハズレ
14	変更	維持	当たり	ハズレ
15	変更	維持	当たり	ハズレ
16	変更	維持	当たり	ハズレ
17	変更	維持	当たり	ハズレ
18	変更	維持	当たり	ハズレ
19	変更	維持	当たり	ハズレ
20	変更	維持	当たり	ハズレ

当たった回数

変更	維持
回	回

「モンティ・ホール問題」の結果集約

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2組	変更	4	5	12	3	2	10	5	7	7	10	5	8	4	3	8	10	7	5	6	7
	維持	3	4	1	0	5	2	7	0	3	2	4	5	4	1	3	3	3	7	3	4
4組	変更	0	5	10	6	5	4	5	3	2	6	5	2	4	6	8	4	5	7	3	4
	維持	3	3	4	3	3	3	2	2	3	3	3	0	3	2	2	1	4	3	4	3
6組	変更	9	5	6	7	10	7	10	6	7	10	6	8	8	1	7	6	8	7	0	4
	維持	1	8	4	3	0	3	3	4	5	3	4	2	2	11	2	2	2	4	5	6

		21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	小計
2組	変更	7	5	2	2	7	8	3	7	5	7	5	7	2	6	8	11	9	5	6	7	247
	維持	1	3	6	2	4	3	7	4	4	3	8	4	3	2	3	2	3	7	4	4	141
4組	変更	6	2	4	5	7	3	3	2	1	7	8	2	6	9	9	5	7	3	3	3	189
	維持	2	2	1	4	3	4	4	4	2	7	5	5	2	6	6	5	3	4	4	4	131
6組	変更	6	5	6	16	10	5	2	5	6	8	4	6	10	7	4	8	6	7	9	11	273
	維持	0	2	2	0	2	5	9	6	2	1	7	5	3	3	7	4	5	4	2	0	143

	変更	維持
総計	709	415
比率	0.63	0.37

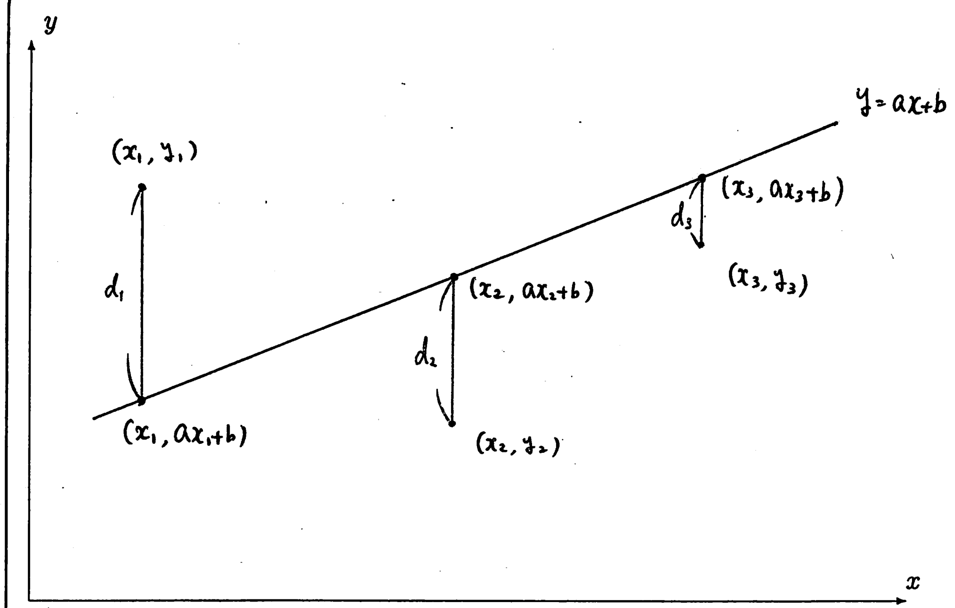
約２：１ですね

こうなる理由を考えてみよう

「回帰直線」のお話

中学校で理科の実験をした際，グラフ用紙にデータを打った後，関係性を見いだすために**どの点からもあまり外れないような直線**を引いたことがあると思います．その直線のことを**回帰直線**と言います．回帰直線を求めるのに使われるのが，あのガウス先生が生み出した**最小二乗法**という考え方です．

最小二乗法の考え方



$d_1^2 + d_2^2 + d_3^2 = \{y_1 - (ax_1+b)\}^2 + \{y_2 - (ax_2+b)\}^2 + \{y_3 - (ax_3+b)\}^2$ を

最小にするような a, b を求める。　d_1, d_2, d_3 は誤差 ← それを小さくしたい！

回帰直線の方程式の係数 a, b を求めるには，大学で習う**「偏微分」**を使うのが一番楽なのですが，**「平方完成」**で**ごりごり**やる方法もあります（裏面参照）．

結論としては，回帰直線 $y = ax+b$ について，$a = \dfrac{S_{xy}}{S_x^2} = r \cdot \dfrac{S_y}{S_x}$　，$b = \bar{y} - a\bar{x}$

回帰直線を求めると，それを利用して予測をすることができるようになります．

３月の平均気温から桜の開花日を予想する

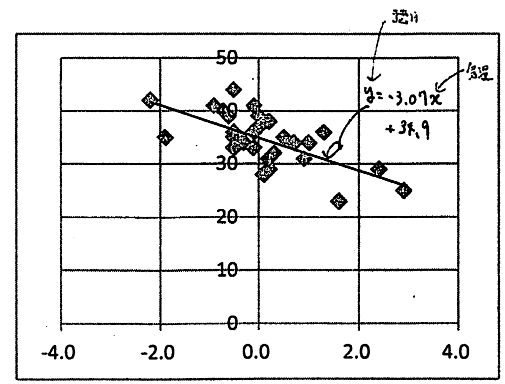
気象庁のWebサイトでは、毎月の平均気温や毎年の桜の開花日などのデータがダウンロードできます。

そこで、１９７０年～２０００年の間の３月の平均気温とその年の桜の開花日のデータ３０年分を基にして回帰直線を求め、２００１年以降の開花日を予想してみました。

※開花日は４月１日からの日数で表しています。

年	平均気温	開花日
1971	0.0	39
1972	0.2	29
1973	-1.9	35
1974	-0.9	41
1975	-0.3	34
1976	0.0	37
1977	-0.1	41
1978	-0.7	40
1979	-0.6	39
1980	-0.5	44
1981	-0.5	35
1982	0.7	34
1983	0.1	28
1984	-2.2	42
1985	-0.1	33
1986	-0.4	35
1987	-0.1	36
1988	-0.1	33
1989	2.4	29
1990	2.9	25
1991	0.3	32
1992	1.0	34
1993	1.3	36
1994	-0.5	36
1995	0.9	31
1996	0.5	35
1997	0.2	31
1998	1.6	23
1999	-0.5	33
2000	0.2	38
平均	0.10	34.60
標準偏差	1.04	4.80
共分散		-3.30
相関係数		-0.66
a		-3.07
b		34.90

年	平均気温	開花日	予想日	誤差
2001	0.2	29	34	+5
2002	2.5	22	27	+5
2003	0.7	31	32	+1
2004	0.7	35	32	-3
2005	0.1	40	34	-6
2006	1.3	38	30	-8
2007	0.9	34	32	-2
2008	3.3	21	24	+3
2009	1.5	31	30	-1
2010	-0.1	37	35	-2
2011	0.7	37	32	-5
2012	0.1	31	34	+3
2013	0.0	43	34	-9
2014	0.5	29	33	+4
2015	3.8	22	23	+1



簡単のために, 2点 $(x_1, y_1), (x_2, y_2)$ の場合を紹介する. 一般の場合も同様に示すことができる. 求める回帰直線を $y = ax + b$ とする.

$$d = {d_1}^2 + {d_2}^2 = \{y_1 - (ax_1 + b)\}^2 + \{y_2 - (ax_2 + b)\}^2$$

を最小にする a, b を求める. なお, $\tilde{d} = \frac{1}{2}d$ で考えて構わない（一般には, $\tilde{d} = \frac{1}{n}d$ で考える）. また, 次の関係式を途中で利用する.

1. 平均値の定義：$\bar{x} = \frac{x_1 + x_2}{2}$, $\bar{y} = \frac{y_1 + y_2}{2}$

2. 分散の定義を変形したもの：${s_x}^2 = \frac{{x_1}^2 + {x_2}^2}{2} - \bar{x}^2$, ${s_y}^2 = \frac{{y_1}^2 + {y_2}^2}{2} - \bar{y}^2$

3. 共分散の定義を変形したもの：$x_1y_1 + x_2y_2 = 2(s_{xy} + \bar{x}\bar{y})$

$$\begin{aligned}
\tilde{d} &= \frac{1}{2}\{({y_1}^2 + a^2{x_1}^2 + b^2 - 2ax_1y_1 + 2abx_1 - 2by_1) \\
&\quad + ({y_2}^2 + a^2{x_2}^2 + b^2 - 2ax_2y_2 + 2abx_2 - 2by_2)\} \\
&= \frac{{y_1}^2 + {y_2}^2}{2} + a^2 \cdot \frac{{x_1}^2 + {x_2}^2}{2} + b^2 - a(x_1y_1 + x_2y_2) + 2ab \cdot \frac{x_1 + x_2}{2} - 2b \cdot \frac{y_1 + y_2}{2} \\
&= {s_y}^2 + \bar{y}^2 + a^2({s_x}^2 + \bar{x}^2) + b^2 - 2a(s_{xy} + \bar{x}\bar{y}) + 2ab\bar{x} - 2b\bar{y} \\
&= \{b^2 - 2(\bar{y} - a\bar{x})b\} + {s_y}^2 + \bar{y}^2 + a^2({s_x}^2 + \bar{x}^2) - 2a(s_{xy} + \bar{x}\bar{y}) \\
&= \{b - (\bar{y} - a\bar{x})\}^2 - (\bar{y} - a\bar{x})^2 + {s_y}^2 + \bar{y}^2 + a^2({s_x}^2 + \bar{x}^2) - 2a(s_{xy} + \bar{x}\bar{y}) \\
&= \{b - (\bar{y} - a\bar{x})\}^2 + {s_x}^2a^2 - 2s_{xy}a + {s_y}^2 \\
&= \{b - (\bar{y} - a\bar{x})\}^2 + {s_x}^2\left(a^2 - 2 \cdot \frac{s_{xy}}{{s_x}^2} \cdot a\right) + {s_y}^2 \\
&= \{b - (\bar{y} - a\bar{x})\}^2 + {s_x}^2\left(a - \frac{s_{xy}}{{s_x}^2}\right)^2 - \frac{{s_{xy}}^2}{{s_x}^2} + {s_y}^2
\end{aligned}$$

よって, $\tilde{d}$ は

$$a = \frac{s_{xy}}{{s_x}^2} = \frac{s_{xy}}{s_xs_y} \cdot \frac{s_y}{s_x} = r \cdot \frac{s_y}{s_x} , \ b = \bar{y} - a\bar{x}$$

のとき, 最小となる.