

『わかりやすい統計学 データサイエンス応用』

<http://www.bayesco.org>

実践力養成問題 各問解答解説

—計算、図・表、解説、実践—

松原 望 解答作成

森本栄一 述、計算、図表作成

リテラシーから技へ
技から習熟へ
習熟からエキスパートへ
エキスパートから名人まで



目次：レベル別と参照章節 一覧

A (リテラシー) からE (名人) へ

※『わかりやすい統計学—データサイエンス基礎』はAレベルです。

1a.1	表読み取り	A	表1.1
1a.2	数字読み取りと要点の計算	C	1.6
1a.3	歴史初の回帰分析	B	表1.7
1a.4	社会的データを読む基本	—	図1.4
1a.5	背景の読解：長文を読む	C	1.5
1a.6	回答法とデータの起源	D	1.3
1a.7	個人の数え方の基準	C	1.3
1b.1	ビットと桁数をRNAで	B	1.8
1b.2	クロス・エントロピーの計算	C	1.11
2.1	暗数の計算	A	表2.2
2.2	よくあるバイアス	B	(4.1)
2.3	サンプル数nは有意性に大きく効く	D	表2.5
2.4	統計分析の主体	D	まえがき
2.5	人間とAIのかかわり	E	(解説)
3a.1	多変量データの要約統計量	A	(基礎編)
3a.2	簡単なクラスター分析	C	3.3
3a.3	簡単な主成分分析例	D	3.2
3b.1	季節調整	B	3.6(2)
3b.2	移動平均をとる	A	図3.4b
3b.3	自己相関係数	C	3.6(4)
3b.4	平均周期	A	3.6(5)
3b.5	AR(1)が定常となる条件	E	3.6(3,5)
3b.6	音程の周波数計算	B	p.70
4.1	幾何学の証明は基本から演繹	C	4.3
4.2	三段論法の例	A	4.3
4.3	帰納推理の例	E	4.4
4.4	統計的センスは単純なケースに	D	4.8
4.5	はずれ値は兆候	E	4.6
4.6	価値の問題は論理からは導けない	E	4.9
5.1	小さい確率の計算	D	5.1, 2
5.2	投資意思決定の期待値の考え方	C	5.1
5.3	組立て乗法で組み合わせの数を計算	E	5.3
5.4	同	E	5.3
5.5	条件付き確率で期待値を計算	D	5.3
5.6	暗号を復号化	D	5.5
5.7	エントロピーのグラフを最大化	D	5.5
6.1	臨床検査は有効か	B	6.1
6.2	eメールのベイズ・フィルター	D	(6.4)
6.3	バレンタイン・チョコ問題	B	6.3
6.4	ベイズ判定関数でアイリス判別	C	6.5
6.5	ベイズの定理とエントロピー計算	D	6.7
7.1	品質管理のパフォーマンス	C	7.1
7.2	リスク中立確率の計算	E	7.3
7.3	ポートフォリオがゼロリスクとなる場合	C	7.2
7.4	標準正規分布の累積分布関数 $\Phi(x)$	B	7.3
7.5	コール・オプションの価格設定の実際	E	7.3

第1章

1a.1 <表読み取り>

男： 23.97 年（令和元年）、24.21 年（同 2 年）

女： 29.17 年（令和元年）、29.46 年（同 2 年）

* まず関心と注意を以って元データを読むことが分析の始まり。[平均余命]と「平均寿命」のちがひ。

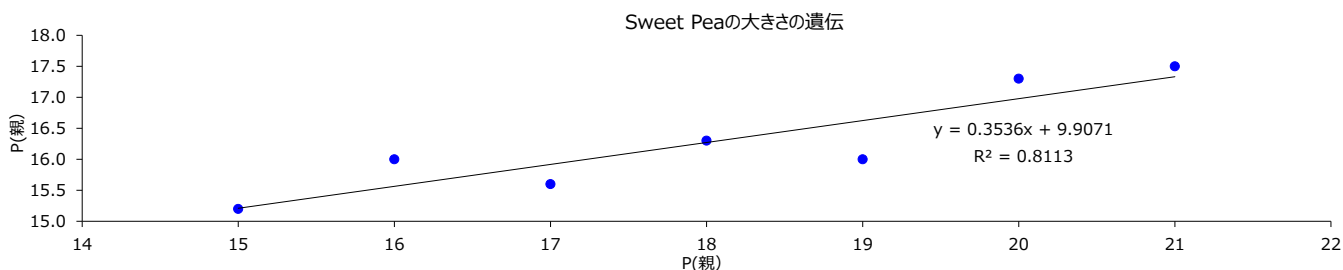
1a.2 <数字読み取りと要点の計算>

簡単でも重要点に関心をもとう。25 年で 2 倍という。式で表すと $x^{25}=2$ で、これを解くため

$$25 \log x = 0.3010 \text{ から } \log x = 0.01204 \quad \therefore x = 1.0281$$

(エクセルで $10^{0.01204}$ で計算) 答 : 2.81%

1a.3 <歴史初の回帰分析>



概要

回帰統計	
重相関 R	0.901
重決定 R2	0.811
補正 R2	0.774
標準誤差	0.403
観測数	7

分散分析表

	自由度	変動	分散	分散比F	有意 確率
回帰	1	3.500	3.500	21.5029	0.00565
残差	5	0.814	0.163		
合計	6	4.314			

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	9.907	1.381	7.174	0.0008	6.357	13.457
P(親)	0.354	0.076	4.637	0.0056	0.158	0.550

残差出力

Case	F予測値	残差
1	15.21	-0.01
2	15.56	0.44
3	15.92	-0.32
4	16.27	0.03
5	16.63	-0.62
6	16.98	0.32
7	17.33	0.17

見るように傾きは 1/3 なので子のばらつきは縮小し、父の平均に戻ってゆく（回帰する）傾向がある。

1a.4 <社会的データを読む基本>

略。社会からデータを、あるいはデータから社会を読むことには、さまざまな立場から非常に多様な「答」がありうるが多い。「正解」はないとするのが適切である。

1a.5 <背景の読解：長文を読む>

「健康食品には副作用がない」と信じることから副作用が見逃され、多量摂取で被害が深刻化する危険があるが、適量についての根拠のデータはない。

* 要旨が述べられている場所は短い。文章読解録力や聞く力は統計学の「始まり」である。

1a.6 <回答法とデータの起源>

(i) 無記入でも「罷免すべきでない」に判定される。

(ii) 無記入でも「罷免すべき」に判定される。

回答法がデータの意味を一変させることがある。

1a.7 <個人の数え方の基準>

(i) 「一株一票」で、「一人一票」の民主主義が適用される場ではない。

(ii) 憲法の本質であり、代議制民主主義の基本原則である。

(iii) 国に対する財産上の貢献をした者だけに参政権があり、基本的人権と相いれない。

1b.1 <ビットと桁数を RNA で>

$4 \times 4 \times 4 = 64$ 通りで、 $2^6 = 64$ で最大 6 ビット

答：6 桁 (ビット)

* DNA を RNA に訂正した

1b.2 <クロス・エントロピーの計算>

符号が 4 通りの場合も 2 通りの場合と同様に、クロス・エントロピーは

$$-(1/2) \log_2(1/4) - (1/4) \log_2(1/4) - (1/8) \log_2(1/4) - (1/8) \log_2(1/4) = 2$$

ちなみに、エントロピーは

$$-(1/2) \log_2(1/2) - (1/4) \log_2(1/4) - (1/8) \log_2(1/8) - (1/8) \log_2(1/8) = 0.5268$$

ここで \log_2 はエクセルでは $\text{LOG}(*, 2)$ で計算する。

第2章

2.1 <暗数の計算>

簡単な計算も、事実の理解に有用である。計算は簡単なほどよい。

いあき： $(145,822 - 20,725) / 20,725 = 6.075$

横領： $(28,448 - 5,743) / 5,743 = 3.954$

住居侵入： $(46,559 - 8,943) / 8,943 = 4.206$

* 表 2.2 の訂正：「暴行障害」を「暴行傷害」と訂正

2.2 <よくあるバイアス>

いじめた経験のある生徒は、たとえ回答が無記名であるとしても提出が自主的であるから、発覚を恐れて回答を提出せず、結局いじめた経験のない回答だけが提出される。統計倫理としては、統

計調査は「捜査」ではなく、調査の目的がいじめの有無なら統計調査の誤用である。「いじめをどう思うか」を尋ねるならかろうじて意味がある。よく言われる「選択バイアス」(4.1)の問題と考えてもよい。

2.3 <サンプル数 n は有意性に大きく効く>

(i)両方のサンプルが同じ母集団からであれば、両サンプルで割合 (%) はおおむね同一と考えられる。このとき、母集団は同一なのに、カイ 2 乗統計量はサンプル数 n だけによって決まる(変わって来る) から結論の信頼性が疑われ、ことに分析者の恣意が入るときは結論の説得力は相当に低い。とりわけ n が研究予算によって決まる場合、説得力は偏に金銭に大きく依存することになりかねないので要注意。「お金をかけて好きな結論を得た」と言われないように。

(ii)サンプル数 n が極めて大きい場合でなければ、サンプルからの相関係数 $r=0.02$ からは母集団相関係数 $\rho=0$ の仮説は有意でなく、関連は統計的には否定される。n が極めて大きければ有意になる可能性があるとしても、全体としてそのような状況自体が結論の説得力を弱める。

2.4 <統計分析の主体>

分析が AI (ことに機械学習) の場合、AI の結論の解釈がその結果の出たプロセスを知らずには可能でないことは多いが (画像認識では結果がすべてである)、今後は、地位あるいは理系文系の出身別にかかわらず、解釈を正しく行う最低限の能力と責任は基本である。これは一般的に「分析」についても言えることである。分析法が新しくなったから、より良い結果が得られるとは限らない。データの良しあし、分析法、結果の読み方能力の 3 拍子揃うことが重要である (解答例)。

2.5 <人間と AI のかわり> 答えやすい順とする。

(ii) 人間性にかかわらない分野で人間の日常行為を機械に置き換えとしても、人間性が侵されたとは考えないし (ウィーナー)、AI にとってかわられる心配もありえない。ただし、「人間性にかかわる」か否かは「人間とは何か」に発展するから、社会や文化、宗教などによりさまざまで簡単ではない。

(i) データ自体にも AI のような働きがあり、分析結果が人間自体を左右する恐れがあるとしても、分析をよく理解した結果であれば、人間が主人公であることに変わりはない。しかし、分析を全く理解できなければ (あるいは明らかでなければ、さらには誤っていれば)、表面的な結果に左右され、自己を見失うことは現代のマスメディア社会の危うさである (解答例)。

第3章

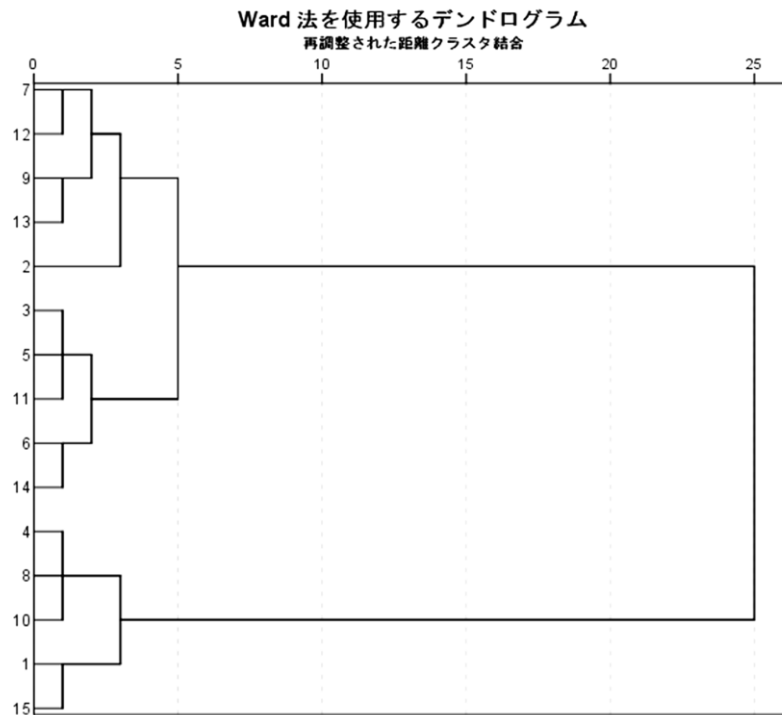
3a.1 <多変量データの要約統計量>

どのような多変量データもまずこれより始め見当をつけるのが常道で、風景画のスケッチと同じである。

	距離	握力	身長	体重
平均	26.20	33.33	156.60	47.40
分散	16.31	48.67	52.26	82.54
標準偏差	4.04	6.98	7.23	9.09

	距離	握力	身長	体重
距離	1			
握力	0.748	1		
身長	0.685	0.530	1	
体重	0.788	0.846	0.694	1

3a.2 <簡単なクラスター分析>



「デンドログラム」(dendrogram (樹形図)) とデシジョン・ツリー (決定樹) を混同しないこと。

3a.3 <簡単な主成分分析例>

SPSS は主成分と各変数の相関係数を出力する。

	第 1 主成分	第 2 主成分
距離	0.910	0.005
握力	0.885	-0.396
身長	0.811	0.564
体重	0.942	-0.118
固有値	3.154	0.489
寄与率(%)	78.855	12.217
累積寄与率(%)	78.855	91.072

主成分負荷量 (上表改変)

	第 1 主成分	第 2 主成分
距離 / $\sqrt{\text{固有値}}$	0.512	0.007
握力 / $\sqrt{\text{固有値}}$	0.498	-0.566
身長 / $\sqrt{\text{固有値}}$	0.456	0.807
体重 / $\sqrt{\text{固有値}}$	0.530	-0.169

主成分分析は表のごとく二乗和 = 1 となる係数として負荷量を出力するが、相関係数とは√固有値倍の違いがある。因子解釈としては、第一主成分（79%）は、大きいこと、強いことの「大きさ因子」として圧倒的、第二主成分（12%）は長さ対重さ・力である。これで主成分分析がほぼ完成する。

3b.1 <季節調整>

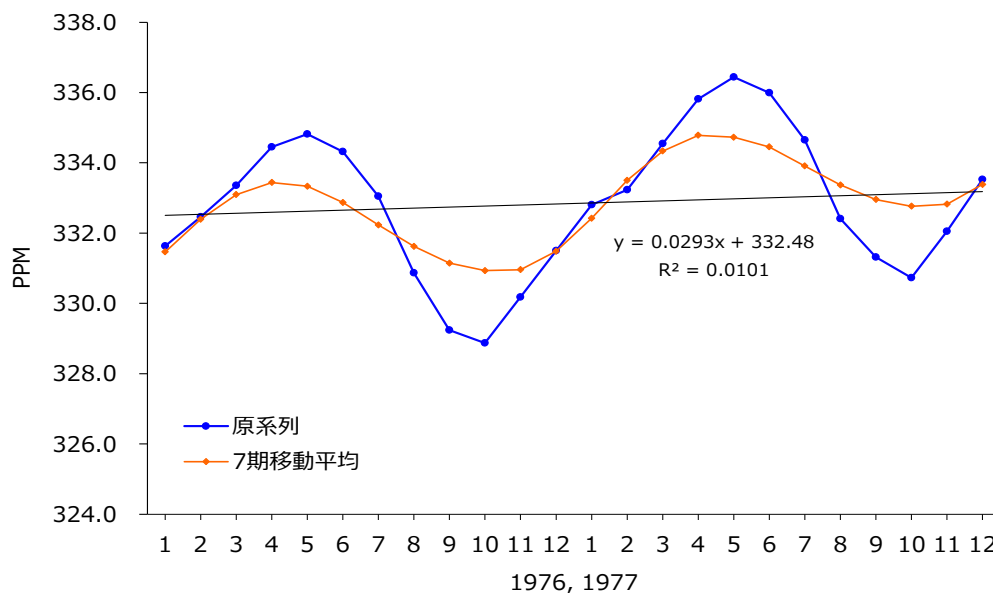
季節性除去は移動平均を用いるが、一般には中央に大きい重みがある。しばしば年換算される。

$$\{ 677,873 + 2 (689,857 + 726,254 + 813,648) + 698,974 \} / 2 = 2,918,183$$

3b.2 <移動平均をとる>

7 期移動平均とは各時点の前後 3 期併せて 7 期の平均。移動平均をとると平滑化されるが、それがトレンド線（傾向線）として使われることもある。もっとも一般的なトレンドは回帰直線である。

大気中CO2濃度原系列、7期移動平均



3b.3 <自己相関係数>（正式には自己相関関数）

ラグが小さければ相関係数は通常の相関係数の式を用いても大差ないから、エクセルで列をスライドし共通部分に対し、コマンド CORREL を用いて、 $r(1)=0.463$, $r(2)=-0.102$

* 正式には自己相関係数の厳密な定義がある。

3b.4 <平均周期>

データに現れた最初、最後の極小期は 1755, 1923 年で、15 周期を含むから 1 平均周期は $(1923-1755)/15=11.2$ (年) これが最も簡便な周期の求め方であるが、あくまで平均周期である。

3b.5 <AR(1)が定常となる条件>

特性方程式 $a^2 - 0.7368a + 0.311 = 0$ を解くと $a = 0.3684 \pm 0.4187i$ 。その絶対値は $|a| = \sqrt{0.3684^2 + 0.4187^2} = 0.5577 < 1$ 。

* 自己回帰モデルだからといって定常とは限らない。計量経済学では後退演算子として扱うので (必ずしもその必要はない)、方程式の次数順序が逆になり条件は > 1 となる。

3b.6 <音程の周波数計算>

ド音の周波数 = $440 \times 3/5 = 264$ Hz だから

(i) ファ音の周波数 = $264 \times 4/3 = 352$ Hz

(ii) ソ音の周波数 = $264 \times 3/2 = 396$ Hz

(iv) ド音の周波数 = $264 \times 2 = 528$ Hz

音波の時系列解析から、音楽も数学の対象で得ることがわかる。

第4章

4.1 <幾何学の証明は基本から演繹>

五角形を ABCDE とし、AC, AD の補助線を入れると、3通りの三角形に対しそれぞれ

$$\angle CAB + \angle B + \angle BCA = 2 \angle R$$

$$\angle CAD + \angle ADC + \angle ACD = 2 \angle R$$

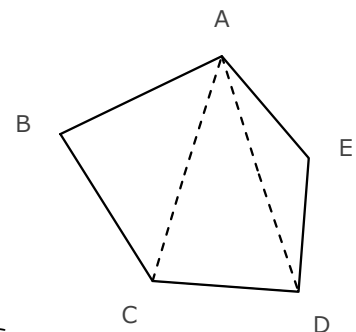
$$\angle DAE + \angle ADE + \angle E = 2 \angle R$$

となる。これらより、まず3式を加えると、3文字の角の和は順次、それぞれ $\angle A$ 、 $\angle D$ 、 $\angle C$ に等しく、よって

$$\angle A + \angle D + \angle C + \angle B + \angle E = 6 \angle R$$

がしたがう。

このように、一般的には数学的「証明」とは演繹論理によるものをいう。



4.2 <三段論法の例> (AAA 式)

大前提 すべての人は人権を持つ (A)。

小前提 すべての日本人は人である (A)。

結論 すべての日本人は人権を持つ (A)。

*大前提、小前提、結論：三段論法の推論では、このように横線を引き、上側を前提、下側を結論という。さらにこここの「すべての…は…である」型の命題を「全称肯定」といい、'A'で表している。参考として、ほかに E (全称否定)、I (特称肯定)、O (特称否定) があり、したがって、三段論法には 64 通りある。実際はさらに 256 通りに分類されるがここでは述べない。これらすべては「論法」であり正しいとは限らないが、正しいか正しくないかのいずれかに分類される。このことは、AI の構成に有益であろう。

参考：坂本百大、坂井秀寿『現代論理学』東海大学出版会

4.3 <帰納推理の例>

人がこれまでに見た「馬」のすべてを集めれば極めて多数で、その限りで馬はすべて四つ足で例外がなかったから、「馬は四つ足」であることは統計的には極めて確からしい。馬のすべてではなく言い切るのには「飛躍」があるが、むしろこの飛躍によってさしあたり「馬は四つ足である」という科学的知識が得られたのであるから、その言い方に十分合理的な理由があり、許される。

4.4 <統計的センスは単純なケースに>

簡単なデータ数字が与えられているとき、与えられ方はさまざまで、そこから意味を引き出す決められた方法も思いつかず、簡単なのに難しい。だが、データには何らかの意味があるのがふつうで、一通りでも推論を思いつくのが統計的センスである。たとえば、次の分析例はどうだろうか。

死者数では 全国：112.5 件で 1 人 都内：235.4 件で 1 人

負傷者数では 全国：1 件あたり 1.25 人 都内：1 件で 1.14 人

「交通事故が死につながるリスクは都内は全国の半分であるが、負傷につながるリスクには都内と全国では大きな差はない。1 を多少越えていることは、運転者のほか巻き込まれた関係者がいることの説明であろう。」

「件数（人数）の都内対全国の比較では

発生件数では 1 : 10.5

死者数では 1 : 21.9

負傷者数では 1 : 11.6

であり、やはり全国の方が死亡のリスクが相対的に高いことを示している」。

* 問は「このデータだけ」から統計的に推論するセンスとを求めている。本題はある大学で出した学期試験である。電卓持ち込み可、配当時間は 90 分 10 題のうちの一題である。

4.5 <はずれ値は兆候>

「はずれ値」（アウト라이어）と考えれば、考えるきっかけになる。統計学では、何か変わった小数の「例外」が起きたと無視はしない。たとえ一例でも、外れ値は重要な意味や背景を持つ場合があるので、さしあたり無視せず慎重に背景を推論すれば問題は防げたかもしれない。

もっとも、データの背景まで調べなくてはならないなら、統計学の問題というより、確率的推論の問題やさらには兆候をとらえる行動科学的な能力（リスク認知力）と考える方が、さらに社会的にも前向きであろう。

4.6 <価値の問題は論理からは導けない>

「そこがいいのだよ」という逆の言明もありうるから、論理的に必然に導かれた結論とするのは適切ではない。日常会話でこういう分別はないから、全体的な混乱しないコミュニケーションの良識が重要となる。

第5章

5.1 <小さい確率の計算>

ロイヤル・ストレート・フラッシュの確率 = $4/2,598,960 = 1.54 \times 10^{-6}$ (100 万分の 1.5)

一等の出る確率 = $7/13,000,000 = 5.4 \times 10^{-7}$ (1000 万分の 5)

ロイヤル・ストレート・フラッシュの確率が一等の出る確率の約 3 倍大きい。非常に小さいので比較が無意味になる(あるいはこれほど小さければ 3 倍は大きな違いではない)とも考えられるが、重要なことはこの程度に小さくとも現実には起こることであり、無視できない。

5.2 <投資意思決定の期待値の考え方>

絵札は 12 枚、それ以外は 40 枚だから期待値は

$$E = \frac{12 \times 35 + 40 \times (-3)}{12 + 40} = 5.76 < 6.00 \quad (\text{ドル})$$

よって、どちらかの目安としては、参加しないほうがよい。ただし、期待値自体が結果となることはない。なお、この式には「確率」は入っていない。

5.3 <組立て乗法で組合わせの数を計算>

					1	1	1	1	1	1	1
					×)	1	1	1	1	1	1
						1	1	1	1	1	1
						1	1	1	1	1	
				1	1	1	1	1	1		
			1	1	1	1	1	1			
		1	1	1	1	1	1				
	1	1	1	1	1	1					
	1	1	1	1	1						
和 =	12	11	10	9	8	7	6	5	4	3	2

このように、係数だけを取り出すための関数は「母関数」といわれる。変数は名目だけであり、t,x など何でもよく、ここでも表示していない。

5.4 <同> 5.3 と同様に

					1	1	1	1
					×)	1	1	1
						1	1	1
						1	1	1
				1	1	1	1	
			1	1	1	1		
		1	1	1	1			
	1	1	1	1				
和 =	8	7	6	5	4	3	2	1 (通り)

								1	2	3	4	3	2	1
							x)	1	2	3	4	3	2	1
								1	2	3	4	3	2	1
								2	4	6	8	6	4	2
						3		6	9	12	9	6	3	
					4	8		12	16	12	8	4		
			3		6	9		12	9	6	3			
		2	4		6	8		6	4	2				
	1	2	3	4	3	2		1						
	1	4	10	20	31	40	44	40	31	20	10	4	1	(通り)
和 =	16	15	14	13	12	11	10	9	8	7	6	5	4	

前問と異なり、母関数の方法だけしか解法を考えられない。

5.5 <条件付き確率で期待値を計算>

条件付き期待値 $E(S|S \geq 13)$ を求める。まず、

$$P(S \geq 13) = 0.0781 + 0.0391 + 0.0156 + 0.0039 = 0.1367$$

よって、 $P(k | S \geq 13)$, $k = 13, 14, 15, 16$ は

$$0.0781/0.1367, 0.0391/0.1367, 0.0156/0.1367, 0.0039/0.1367$$

から、下表となる。

13	14	15	16
0.5714	0.2857	0.1143	0.0286

条件付き期待値は

$$E(S | S \geq 13) = 13 \cdot 0.5714 + 14 \cdot 0.2857 + 15 \cdot 0.1143 + 16 \cdot 0.0286 = 13.60$$

条件付き確率はよく知られるが、条件付き期待値は慣れない人も多いだろう。

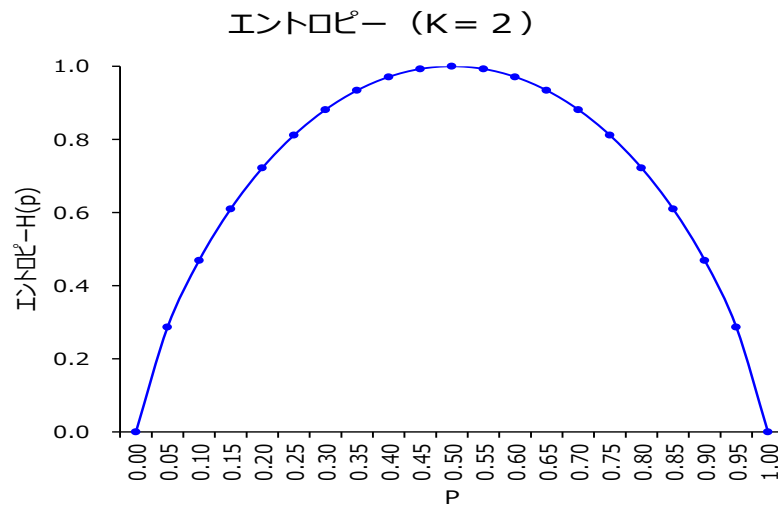
5.6 <暗号を復号化>

先頭から、110 111 00 110 と区切られるので“DEAD”と復号化される

*このように区切る以外にない。さらに、先頭から区切る方法で復号化される。

5.7 <エントロピーのグラフを最大化>

下図、 $p=0.5$ で最大 1 ビットとなる。予測結果があいまいで、実質は予測になっていない。エントロピーの図は珍しいが、放物線 $4x(1-x)$ に似る。なお、計算で、 $0 \log 0 = 0$ と約束する。



第6章

6.1 <臨床検査は有効か>

$$\text{感度} = \frac{70}{70 + 40} = 0.636, \quad \text{特異度} = \frac{20}{20 + 200} = 0.099$$

$$\text{陽性反応正診率} = \frac{70}{70 + 20} = 0.778, \quad \text{陰性反応正診率} = \frac{200}{40 + 200} = 0.833$$

感度、特異度の実際例：矢野栄二、小林康毅、山岡和枝『EBM 健康診断』医学書院など

6.2 <eメールのベイズ・フィルター>

ベイズ統計学がIT、AIへ適用された最初とされているが、ベイズの定理の要点を内容とする点でベストな例である。「通常メール」の事後確率を求める。メールの事前確率が**6対4**に注目して

$$(i) \text{「無料」が見出されたとき} \quad \frac{0.6 \times 0.3}{0.6 \times 0.3 + 0.4 \times 0.4} = \frac{9}{17} = 0.529$$

次に、この事後確率が事前確率に更新（上書き）されるから

$$(ii) \text{「当選」が見出されたとき} \quad \frac{\binom{9}{17} \times 0.7}{\binom{9}{17} \times 0.1 + \binom{8}{17} \times 0.6} = \frac{9}{9 + 48} = 0.158$$

データの統計的確率では「通常」が「迷惑」に対し1/6しかないので、「通常」の事後確率は激減した。(0.529→0.158)

判定：判定の境界確率次第であるが、0.8とすると(i)の段階で「否」、(ii)でそれが確実となる。また0.5とするときには、(i)ではかろうじて「合」であるが、再検すると(ii)で「否」となる。結局、常識的決定ルールでは「否」である。確率が決定に役立つ点でもいい例である。

6.3 <バレンタイン・チョコ問題> 6.2同様、事後確率が次段階の事前確率に更新（上書き）されるところがポイントである。よって

$$\frac{0.752 \times 0.65}{0.752 \times 0.65 + 0.248 \times 0.35} = \frac{0.489}{0.489 + 0.087} = 0.862$$

85%を超えたが、それをどう考えるか、決定は本人次第である（その意味では前問と同様）。

6.4 <ベイズ判定関数でアイリス判別> ストレートに計算すると

$$F_{\text{バージニカ・ベルシカラ}} = -40.1774 < 0, \quad F_{\text{バージニカ・セトーサ}} = -87.4353 < 0$$

$$F_{\text{セトーサ・ベルシカラ}} = -47.2579 < 0$$

正、負の符号から、あるいは判別領域の図からセトーサと判定。機械学習（SVM）と似る。

6.5 <ベイズの定理とエントロピー計算> 事前確率は指定されておらず、等確率 1/3 とする。

(i) 赤玉の場合：

$$P(A_1|\text{赤}) = \frac{(1/3)(4/3)}{(1/3)(4/3) + (1/3)(1/2) + (1/3)(1/3)} = \frac{9}{19}$$

$$P(A_2|\text{赤}) = \frac{(1/3)(1/2)}{(1/3)(3/4) + (1/3)(1/2) + (1/3)(1/3)} = \frac{6}{19}$$

$$P(A_3|\text{赤}) = \frac{4}{19}$$

(ii) 白玉の場合：(3/4, 1/2, 1/3) を(1/4, 1/2, 2/3) に変えて、同様の計算から

$$P(A_1|\text{白}) = \frac{3}{17}, \quad P(A_2|\text{白}) = \frac{6}{17}, \quad P(A_3|\text{白}) = \frac{8}{17}$$

(iii) 事前確率 (1/3, 1/3, 1/3) のエントロピー (ビット)

$$-(1/3) \log_2(1/3) - (1/3) \log_2(1/3) - (1/3) \log_2(1/3) = 1.5850 \quad \dots \textcircled{1}$$

事後確率(9/19, 6/19, 4/19)のエントロピー (ビット)

$$-(9/19) \log_2(9/19) - (6/19) \log_2(6/19) - (4/19) \log_2(4/19) = 0.5625 \dots \textcircled{2}$$

事後確率(3/17, 6/17, 8/17)のエントロピー (ビット)

$$-(3/17) \log_2(3/17) - (6/17) \log_2(6/17) - (8/17) \log_2(8/17) = 0.4602 \quad \dots \textcircled{3}$$

よって、獲得情報量は① > ②, ① > ③から

$$\text{赤玉の場合} : 1.5850 - 0.5625 = 1.0225,$$

$$\text{白玉の場合} : 1.5850 - 0.4602 = 1.1248$$

結論として、白玉の方が獲得情報量が多い。計算自体はむずかしくないが、目に見えない「情報」のこのような計算ができることが「情報」について学ぶ真の意義である。

第7章

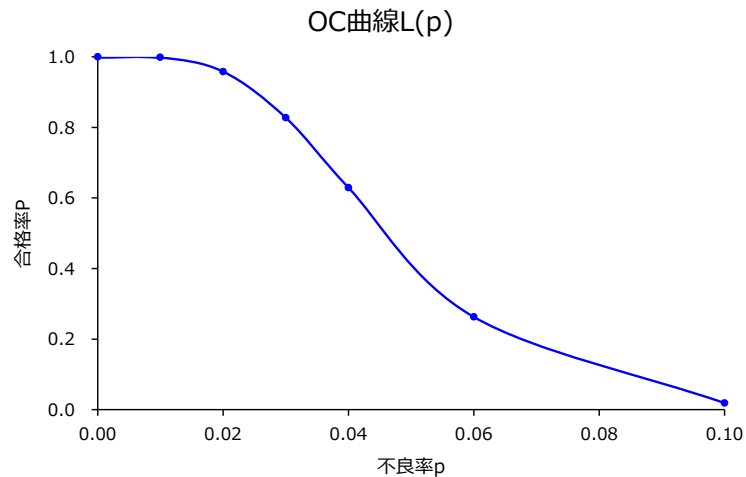
7.1 <品質管理のパフォーマンス> 読むべき該当する数字がないのでxからyを決める線形内挿の式

$$y = \frac{y_1 - y_2}{x_1 - x_2} (x - x_2) + y_1$$

で P, p を x, y として

$$\frac{0.03 - 0.04}{0.8276 - 0.6295} \times (0.7 - 0.8276) + 0.03 = 0.0364 \quad \text{答 } 0.0364 \text{ (3.64\%)}$$

* 参考として、「作用特性曲線」(OC curve) は下図である。



7.2 <リスク中立確率の計算>

1000ドル単位で

$$60p + 42(1 - p) = 50e^{3 \times 0.03}$$

を解くと、

$$18p = 50e^{0.09} - 42$$

これより、リスク中立確率は

$$p = \frac{50e^{0.09} - 42}{18} = 0.706$$

リスク中立確率は、このように、「世の中」をシミュレートした確率である。

7.3 <ポートフォリオがゼロリスクとなる場合> ポートフォリオの分散の式

$$\sigma^2 = x^2\sigma_A^2 + (1 - x)^2\sigma_B^2 + 2\rho\sigma_A\sigma_B$$

において理想的に $\rho = -1$ とすると

$$\sigma^2 = \{x\sigma_A - (1 - x)\sigma_B\}^2$$

となるから

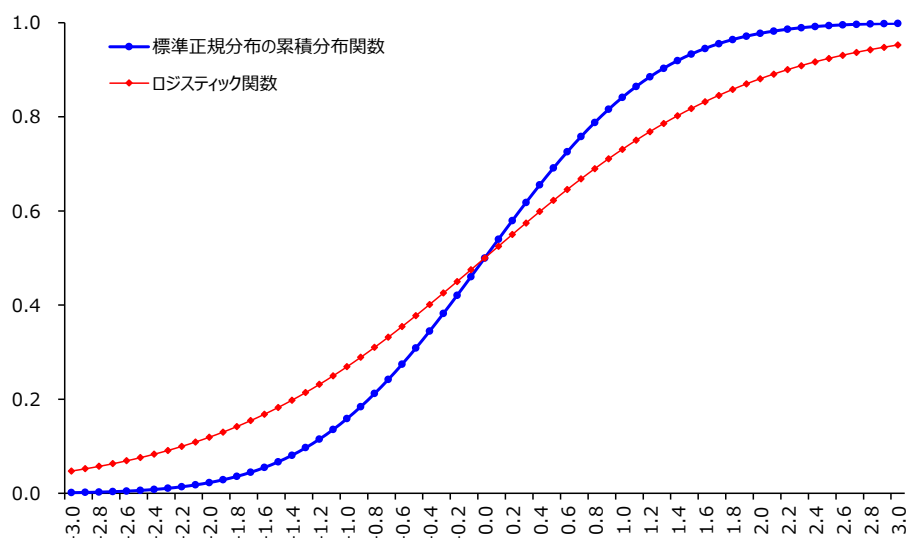
$$x\sigma_A = (1 - x)\sigma_B, \text{ すなわち } x:(1 - x) = \sigma_B:\sigma_A$$

と決めることで、ポートフォリオは無リスク $\sigma = 0$ となる。

7.4 <標準正規分布の累積分布関数 $\Phi(x)$ >

長い語だが正式用語である。金融工学では $N(x)$ とあらわされている。

標準正規分布の累積分布関数
(シグモイド型)



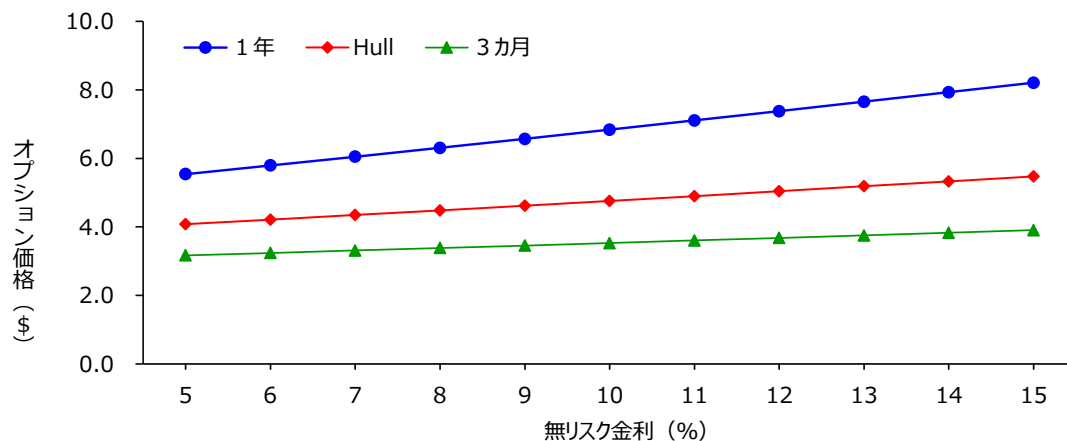
*シグモイド型：「擬S型」と訳す。図で前者はプロビット分析、後者はロジット分析で使われる。

7.5 <コール・オプションの価格設定の実際>

- ①本文通り、4.759
- ②(i) (ii)下表参照
- ③下図参照 (Hull = 半年)

金利 (%)	5	6	7	8	9	10	11	12	13	14	15
1年	5.543	5.794	6.049	6.308	6.571	6.837	7.107	7.379	7.654	7.931	8.211
半年	4.081	4.213	4.347	4.483	4.620	4.759	4.900	5.042	5.185	5.330	5.476
3カ月	3.169	3.240	3.311	3.383	3.456	3.529	3.603	3.678	3.753	3.828	3.904

コール・オプションの価格設定
(対無リスク金利)



第8章、第9章の実践力養成問題も企画作成中です。ご期待ください。