付録 E

解答と解説

■第 1 章　データの整理と表現

問題 1–1

▼平均

$$\mu = \frac{1}{100}(43.6 + 45.2 + \cdots + 64.6) = 54.461$$

答え： 54.46

▼分散, 標準偏差　次のように式 (1.6) を使うと，平均からの偏差をいちいち求める必要がなくなり，計算の手間が省ける．

$$\sigma^2 = \frac{1}{100}(43.6^2 + 45.2^2 + \cdots + 64.6^2) - 54.461^2 = 17.841$$

答え： 分散 17.84, 標準偏差 4.224

上の計算では計算途中の数値の桁数を大きく取っていて，最後の答えのときだけ指定された有効数字に丸めていることに注意．有効数字と数値の丸め誤差については付録 C を参照のこと．

分散の定義式 (1.4) をそのまま使うと，次のようになる．結果は同じになるはずだが，若干誤差が出やすい．

$$\sigma^2 = \frac{1}{100}\left\{(43.6 - 54.461)^2 + (45.2 - 54.461)^2 + \cdots + (64.6 - 54.461)^2\right\} = 17.84$$

$$\sigma = \sqrt{17.84} = 4.2237\ldots$$

問題 1–2

0 は $n(1-p)$ 個，0 は np 個ある．つまりデータ列は次のようになるわけだ．

$$\overbrace{0, 0, \ldots, 0}^{n(1-p)}, \overbrace{1, 1, \ldots, 1}^{np}$$

式 (1.2) を使って平均を求めると，次のようになる．

$$\mu = \frac{1}{n}(\overbrace{0+0+\ldots+0}^{n(1-p)} + \overbrace{1+1+\ldots+1}^{np}) = \frac{np}{n} = p$$

あるいはこのデータを 階級値 0, 1 だけをもつ度数分布とみなすと，17 ページにあるように，平均は階級値とその割合の積の和に等しい．0 の割合は $1-p$, 1 の割合は p であるから答えはとても簡単である．

$$\mu = 0 \times (1-p) + 1 \times p = p$$

分散を求めるには，2 乗の平均から平均の 2 乗を引くやり方が便利だ．

$$\sigma^2 = \frac{1}{n}(\overbrace{0^2+0^2+\ldots+0^2}^{n(1-p)} + \overbrace{1^2+1^2+\ldots+1^2}^{np}) - \mu^2 = p - p^2$$

標準偏差は正の平方根をとって $\sqrt{p-p^2}$ となる．

問題 1–3

体重の 2 乗の和を求める．式 (1.7) にならって，次のように書いてみよう．

$$2\,\text{乗の和} = 44.0^2 + \overbrace{46.0^2 + 46.0^2 + 46.0^2}^{3} + \overbrace{48.0^2 + 48.0^2 + \cdots + 48.0^2}^{6} + \cdots$$

したがって 2 乗の平均は，次のようにして求めることができる．

$$\frac{2\,\text{乗の和}}{\text{総人数}} = \frac{44.0^2 \times 1 + 46.0^2 \times 3 + \ldots + 64.0^2 \times 2}{1+3+\ldots+2} = \frac{298308}{100} = 2983.08$$

これから先に求めておいた平均の 2 乗を引いてやれば，分散が求められる．

$$2983.08 - 2965.89 = 17.19$$

問題 1–4

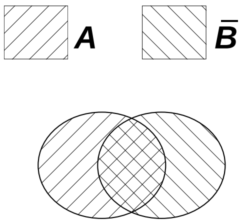
度数分布表から計算すると，平均は 7.3 歳，メジアンは 5 歳となる．幼稚園に 7.3 歳の人がいることは考えられないので，この平均値は実態を反映しておらず，少数の大人（ゆか先生，あゆみ先生，お母さん先生，園長先生，おばあちゃん先生，おじいちゃん先生と

いうところかも）の年齢に引っ張られている．一方で，むしろメジアンのほうが 5 歳を中心とした構成になっていることをよく表現していることになる．

■第 2 章　初等的な確率論

問題 2–1

下図で A と $\overline{B}$ のいずれにも含まれる領域が $A - B$ であることは一目瞭然．



問題 2–2

ベン図を見ると，A, B の排他的論理和は $A - B$ と $B - A$ の和集合となっている．従って，次のように表される．

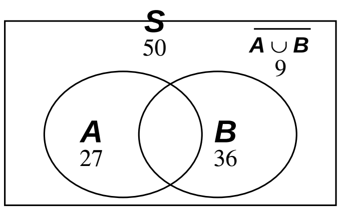
$$(A - B) \cup (B - A)$$

さらに問題 2–1 の式を使えば次のようになる．

$$(A \cap \bar{B}) \cup (\bar{A} \cap B)$$

問題 2–3

1. 図のようになる。



2. 全体集合 S の要素数は 50 である．S に含まれる要素のうち A でも B でもない要素の集合，つまり $\overline{A} \cap \overline{B} = \overline{A \cup B}$ の要素の数は 9 なのだから，$A \cup B$ の要素数は 41．つまり $27 + 36 = 63$ は重なりを含めた数で，真の合計は 41 なのだから，$A \cap B$ に含まれる人数は $63 - 41 = 22$ となる．

問題 2–4

A を感染者であるという事象，B を非感染者である事象とすると，A と B は互いに排反で，かつ標本空間を尽くしている．E を検査で陽性になる事象としよう．問題は $P(A|E)$ を求めることである．

題意より，$P(E|B)=0.015$, $P(E|A)=1-0.005$, $P(A)=0.02$, $P(B)=1-P(A)=0.98$ であるから，

$$P(A|E)=\frac{P(A)P(E|A)}{P(A)P(E|A)+P(B)P(E|B)}=\frac{0.02\times 0.995}{0.02\times 0.995+0.98\times 0.015}$$
$$=0.575\ldots$$

よって約 58%．この段階では本当に感染しているかどうかは半々に近いというわけだ．

問題 2–5

連続して同じ数が出現する確率は 1/10 だから，0 から 9 の数字をランダムに 50 書いたとすると，5 回程度は現れる．3 個連続する確率も 1/2 だから，あってもおかしくはない．

問題 2–6

A を黒である事象，B を偶数 (2,4,6,8,10) である事象とする．ただし J, Q, K のカードは数字札とはみなさないこととする．したがって $P(A)=\frac{1}{2}, P(B)=\frac{5}{13}$．一方，黒でありかつ偶数であるという事象 $A\cap B$ は，スペード，クラブの 2,4,6,8,10 の 10 枚のカードの出現に対応する．よって $A\cap B=\frac{10}{52}$．結局

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{1}{2}+\frac{5}{13}-\frac{10}{52}=\frac{36}{52}=\frac{9}{13}$$

問題 2–7

ここでは閏年を考慮しないものとする．クラスに n 人いて，誕生日がだれも一致しない確率を p_n とする．すると一致する人がいる確率は $1-p_n$ で計算できる．

いま，1 人いるところにもう 1 人が来たとしよう．最初の 1 人によって 365 日のうち 1 日は誕生日に当たっているから，2 人目がそれと一致しない確率 p_2 は 364/365 となる．次にもう一人が来たら，すでに 2 人がいて誕生日が一致していないのだから，3 人目の人も一致しない確率は $p_3=363/365$．これを続ければ n 人の誕生日が一致しない確率は次のようになる．

$$p_n=\frac{364}{365}\times\frac{363}{365}\times\frac{362}{365}\times\cdots\frac{365-n+1}{365}$$

電卓などを使って実際に計算していくと，$n = 20,21,22,23$ のときの値は $0.589, 0.556, 0.524, 0.493$ となるので，23 人の時に誰か誕生日が一致する人がいる確率は $1-0.493=0.507$ となって 1/2 を超える．

ちなみにこの答の人数は，たいていの人が常識的に考えるよりもかなり小さくて意外な感じを与える．そこで**誕生日のパラドックス**」と呼ばれることがある．

■第 4 章　二項分布

問題 4–1

二項分布が成立するので，式 (4.1) で $p = 2/3, n = 8$ として計算する．関西出身が 4 人未満だから $x = 0, 1, 2, 3$ について確率を求めて足し合わせればよい．

$$
{}_8C_0\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^8 +{}_8C_1\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^7 +{}_8C_2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^6 +{}_8C_3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^5 \\
= 0.088
$$

問題 4–2

この問題は，3 通りの事象があって，6 回の試行でそれぞれが 2 回ずつ実現する確率とみることができるので，多項分布の式が使える．式 (4.4) で，$n = 6, x_1 = 2, x_2 = 2, x_3 = 3, p_1 = 1/13, p_2 = 1/3, p_3 = 1/3$ とおいて，

$$
f(2, 2, 2) = \frac{6!}{2!2!2!} \times \left(\frac{1}{3}\right)^6 = \frac{10}{81}
$$

と計算すればよい．

問題 4–3

このように比較的小さい期待値で事象が発生する現象にはポアソン分布が有効である．まずある週が無欠席になる確率 p を求めよう．式 (4.5) に $\mu = 2.5, x = 0$ を代入して，

$$
p = \frac{2.5^0}{0!}e^{-2.5} = e^{-2.5} = 0.0821.
$$

求めるのは $p^2 = 0.0067$ となる．

■第 5 章　正規分布

問題 5–1

正規分布表から，$\Phi(z) = 0.75$ となるのは $z = 0.67$ である．一方，データから計算して $\mu = 54.46,\ \sigma = 4.22$ となるから，$\mu \pm 0.67 \times 4.22 = 51.6, 57.2$ を求めればよい．この範囲に入る人数を実際に数えてみると 49 名となっているから，予測は正しいことが分かる．

問題 5–2

偏差値の定義から，偏差値 58 に相当する点数は，標準正規分布で次の z に相当することになる．

$$z = \frac{58-50}{10} = 0.8$$

正規分布表から $\Phi(z) = 0.788$ が求められるので，この人の上位にいる人の数は割合にして $1 - 0.788 = 0.212$．したがって $12000 \times 0.212 = 2544$ となるから，2540 番目ぐらいに位置することになる．

問題 5–3

二項分布では $\mu = np$, $\sigma = \sqrt{np(1-p)}$ である．従ってこの集団については，$\mu = 9.6$, $\sigma = \sqrt{24 \cdot 0.4 \cdot 0.6} = 2.4$ となる．一方，この二項分布を正規分布とみなすとき，9 人以上 12 人以下は $z_1 = 8.5$ から $z_2 = 12.5$ までの区間とみなされる（半整数近似）．そこで $z_1 = (8.5-9.6)/2.4 = -0.458$ $z_2 = (12.5-9.6)/2.4 = 1.208$ であるから，正規分布表から $\Phi(0.46) = 0.6772, \Phi(1.21) = 0.8869$ を探す．z_1 と z_2 に挟まれる区間の面積は $(\Phi(0.46) + \Phi(1.21) - 1) = 0.6772 + 0.8869 - 1 = 0.5641 \approx 0.56$ とすればよい．ここでは $\boldsymbol{\Phi(1.21) - \Phi(0.46)}$ **としてしまう誤り**を犯しがちなので注意．

問題 5–4

全数が n, 確率 p がちょうど 1/2 であるから，女（男でもよい）の子の数の期待値は $\mu = np = 40000 \times \frac{1}{2} = 20000$ となる．また，標準偏差は $\sigma = \sqrt{np(1-p)} = 100$．女の子の数 x が全体の 51% であったとすると $x = 40000 \times 0.51 = 20400$ であるから，標準化変換によって

$$z = \frac{x-\mu}{\sigma} = 4$$

となる．同様にして 49% からは $z = -4$ が得られる．正規分布表を調べると，$\Phi(4) = 0.9999683$ であるから，女の子の数が $-4 < z < 4$ の範囲から外れる確率は $(1 - 0.9999683) \times 2 = 6.3 \times 10^{-5}$ となる．

このように，全体の数が大きい場合には，事象が実際に実現する回数は数学的な期待値に近い値になることが確実になる．これは確率における**大数 (たいすう) の法則**と呼ばれている．

なお，実際の男女の出生比率は 1 : 1 ではなく，男の子のほうがやや多い．

問題 5–5

1. 0.71, 2. 0.98

この問題では，400 人のときには半整数補正の効果が 3% 程度ある．2000 人の時にはほとんど影響しない．

問題 5–6

区間 $[0, 1]$ で一様連続分布する確率変数の平均は 1/2，分散は 1/12 であるから (5.1.6

節参照)，この分布に従う 12 個の変数の和を S とすると，S の平均は $1/2 \times 12 = 6$, 分散は $1/12 \times 12 = 1$ となる．その理由については 3.5 節参照のこと．

したがって，$S-6$ という値は平均が 0, 標準偏差が 1 の連続分布になる．一方，多数の独立な確率変数を足し合わせると正規分布に近づくことが中心極限定理で保証されているから，$S-6$ は $N[0,1]$ の標準正規分布に従うことになる．

■第 6 章　無作為抽出と標本分布

問題 6–1

このように整然と区画を選ぶと，もともと畑全体に存在していた何らかの傾向，たとえば縁にある区画が余計に選び出されるので作物の生育環境の偏った区画を選んでしまうといった種類の偏りを反映してしまう可能性がある．すべての区画に 1 から 100 までの番号を振っておいて，乱数を使って 20 の区画を選び出すのが正しい方法である．このさい，たとえば隣同士が選び出された場合にそれは避けるといった，事後の人為的な調整も行わない方がよい．

問題 6–2

ある本を買う人は，その本の内容に関心があり，かつそれを肯定的にとらえる傾向があると考えられる．また，その本を読むことによって著者の主張を受け入れる可能性も高い．したがって，読者アンケートによる調査は，信頼できる客観的な調査とはいえない．

ちなみに，血液型が性格と関係があるという説に対しては何人かの心理学者による調査があるが，それを肯定する結果はほとんどなく，また科学的にも関連性があるという根拠はない．ただし，最近の韓国での研究で，B 型の男性について有意な差がみられるという結果があった．これについては韓国で血液型ブームが起こり，歌や映画で B 型が嫌われる風潮を扱ったりしたことによって，思い込みによる回答が無視できないほどにまでなったことによるものであると考えられている．

問題 6–3

題意から，母平均は $\mu = 54.2$, 母分散は $\sigma = 0.22^2 = 0.0484$ である．標本のサイズは $n = 10$ なので，式 (6.3) より，標本平均 $\bar{X}$ の平均値は 54.2, 式 (6.4) より標本平均の分散は $V[\bar{X}] = \sigma^2/n = 0.00484$ だから，その標準偏差は平方根をとって 0.0696 となる．標本平均の 54.1, 54.3 をこれらを使ってそれぞれ標準化変換すると，

$$z_1 = \frac{54.1 - 54.2}{0.0696} = -1.44, z_2 = \frac{54.3 - 54.2}{0.0696} = 1.44$$

となる．正規分布表から $\Phi(1.44) = 0.92506$ を読み取って，この 2 つの値の間に入る確率を求めると，確率として 0.85 が得られる．

問題 6–4

24.4 g^2 (解法はテキスト参照のこと)

■第 7 章 推定

問題 7–1

この問題には二種類の解法が考えられる．

1. **支持を 1, 不支持を 0 として標本平均と標本分散を求めて計算する方法：**
   支持を 1, 不支持を 0 とすると, 100 人の標本平均は $\bar{X} = \frac{1}{100}(1\times 30+0\times 70) = 0.3$
   標本分散は $s^2 = \frac{1}{100}((1-0.3)^2 \times 30 + (0-0.3)^2 \times 70) = 0.21$。
   したがって母分散の不偏推定値は $\sigma^2 = \frac{n}{n-1}s^2 = 0.2121$ より，$\sigma = 0.4605$。これより $\bar{X}$ の分布の標準偏差は $\sigma/\sqrt{100} = 0.04605$ となる。
   求めるのは平均値を挟む 95% の信頼区間であるから，97.5% 点を使って，$0.3 \pm 0.04605 \times 1.960$ を求めて，内閣支持率は 0.21 〜 0.39 と推定される．
2. **支持者の数に関する二項分布を使う方法：**
   標本における支持率を p とし，$n = 100$ 人のうち支持している人の数を x とすると，x は二項分布に従うから標本平均は $\bar{X} = np = 30$, 標本分散は $s^2 = np(1-p) = 100 \times 0.3 \times 0.7 = 21$ である．これから母分散 σ^2 の不偏推定値は $\sigma^2 = \frac{n}{n-1}s^2 = 21.21$ となる．そこで 95% 信頼区間の幅は
   $$\sigma \times 1.960 = \sqrt{21.21} \times 1.960 = 9.027$$
   つまり，30 ± 9.027 が信頼区間に入るので，人数の 100 で割って内閣支持率は 0.21 〜 0.39 と推定される．

問題 7–2

標本が小さいので $t-$ 分布を使う．標本の大きさは $n = 12$ であるから，自由度は $\nu = n-1 = 11$, 99% 信頼区間を計算するには $\alpha = 0.005$ として，3.106 を表から読み取る．ここで Student の変数 T を計算すると，

$$T = \frac{\sqrt{n-1}(x-\mu)}{s} = \frac{\sqrt{11}(16-\mu)}{4}.$$

ここから $T = 3.106$ および $T = -3.106$ として 推定値 μ を求めればよい．結果は 12.3 〜19.7 となる．

問題 7–3

この場合には大標本の取り扱いをしてよい．したがって母平均の推定は 7.3.2 節の議論をもとに式 (7.8) と同様にして考えればよい．99% 信頼区間は次のようになるので，

$$\left[16 - 2.576 \times \frac{4}{\sqrt{80-1}}, 16 + 2.576 \times \frac{4}{\sqrt{80-1}}\right]$$

計算して，14.8〜17.2 を得る．

この結果を問題 7–2 の結果と比べると，推定区間の幅はかなり小さくなっていることがわかる．これは当然で，標本が小さいほど，データから推定できる幅は広く，つまりあいまいにならざるを得ないということを反映しているのである．

■第 8 章　仮説と検定

問題 8–1

実の総数は $n = 144$．このとき 1 代目が黄色と白それぞれの純系種であったとすると，メンデルの法則に従って黄色の実ができる確率は $p = 3/4$ である．したがって黄色の実の数の期待値は $np = 108$，標準偏差は $\sigma = \sqrt{np(1-p)} = 5.196$ となる．実際に得られたのは 119 であるから，これを標準化変換してみると，

$$z = \frac{119 - 108}{5.196} = 2.11$$

となる．この値を正規分布のパーセント点と比較すると，危険率 5% であれば 97.5% 点（あるいは $\alpha = 0.025$) の 1.960 との比較になるので，仮説は棄却される．つまり，1 代目は純系でなかったという判断になる．もちろんこの判断は 20 回に 1 回程度以内の間違いを含むことを許容しているのである．

一方，危険率 1% であれば 99.5% 点（あるいは $\alpha = 0.005$) の 2.576 との比較になるので，仮説は棄却されない．つまり，1 代目は純系であったという仮説を棄却できない．ただしこの判断は，上の判断よりも厳しい基準を設けてずっと用心深く振る舞っているのであり，2 つの判断の間に矛盾はない．

問題 8–2

1. 母平均 $\mu = 12.6$ g, 母標準偏差 $\sigma = 1.9$ g の母集団から 12 個抽出を行ったときの標本平均の分布は，(標本平均の) 平均が 12.6 g,(標本平均の) 標準偏差が $1.9/\sqrt{12}$ の正規分布になる． これを使って与えられたデータから得られた実際の標本平均 $\bar{X} = 12.725$ を標準化変換すると，0.228 となる．正規分布で $\alpha = 0.025$ となる点は 1.960 であるから，標本平均の値は棄却域に落ちない．つまり，仮説は棄却されない．

2. 式 (8.2) を使って Z を計算してみる．

$$Z = \frac{1}{\sigma^2}\left((11.78 - \bar{X})^2 + (12.92 - \bar{X})^2 + \cdots\right) = 25.62$$

この Z が母集団（現存種）から抽出されたデータによる者であるとした場合，自由度 $\nu = 12 - 1 = 11$ の χ^2 分布に従うことになるはずである．そこで危険率 5%(両側検定なので $\alpha = 0.025$) の棄却域を表から調べると 21.920 である．つまり Z の値は棄却域に落ちるので，仮説は棄却される．

これらの結果をまとめると，遺跡から出た貝は，質量は大体現存種とあっているのだが，個体差がかなり大きいという特徴があるので異なる種なのかも知れない．

問題 8–3

式 (8.3) に従って X を計算すると 1.78 が得られる．一方この値は自由度 $(2-1)(2-1) = 1$ の χ^2 に従う．一方付録の χ^2 分布表で $\nu = 1, \alpha = 0.005$ の値を調べると 7.879 であるから，X は 1% の危険率で棄却域の内側にある．したがって，独立であるという仮説は棄却できない．具体的にいうと，この病気にかかることと飲酒の習慣の間に相関があるとはいえないことになる．

■第 9 章　相関と線形回帰

問題 9–1

1. $\sigma_x^2 = 105.32, \quad \sigma_y^2 = 7.801, \quad \sigma_{xy}^2 = 24.40$
2. $\rho_{xy} = 0.851$
3. $a = 0.23, \quad b = 19.0$
4. $T = 3.98$ となる．t-分布を調べると，自由度 $8 - 2 = 6$ のときに $\alpha = 0.005$ の点は 3.707 である．したがって T は棄却域にあるので，仮説は棄却される．つまり，卵の質量と黄身の質量は独立ではないと検定できる．