

『入門 実践する統計学』の解答例

藪友良『入門 実践する統計学』（東洋経済新報社、2012年）の解答例です。練習問題に取り組むことで、統計学の理解が深まりますので、ぜひすべての問題を解いてみてください。

第1章の答え

1. ①連続変数、②離散変数、③連続変数、④連続変数となります。また、①横断面、②横断面、③時系列、④パネルデータとなります¹。
2. 1)サンプルサイズは1000となります。2)この大学は母集団を代表していない可能性があります。全国の大学生から無作為抽出する必要があります。
3. 銀座にいる人は、比較的高所得者が多いと考えられます。また、年収の高い人の方がアンケートに答えてくれやすい、低い人は答えにくい傾向もあるかもしれません。
4. 合格する自信のある人が事前予約していると考えられます。部屋探しのコスト（部屋を探す手間）を支払ってまで事前予約する人は、自分が合格するという自信を持っているのです。
5. 監督官庁の聞き取り調査は、その企業が不正行為を隠そうとするバイアスがかかってしまうものです。たしかに回収率は高くなりますが、本当のことを答えていない可能性があります。
6. 太郎君は見栄をはって、正直に答えなかった可能性があります。
7. 対面調査の利点は、明らかな嘘を見抜くことができること、回答者の記入漏れなどもなくなることです²。欠点は、面接員の特性が相手に何らかの影響を与えること（問6の太郎君の行動が該当します）、面接員の手抜きが生

¹所得は1円単位であり厳密には連続変数ではありませんが、そのとりうる値は無数にあり連続変数として扱われるのが普通です。

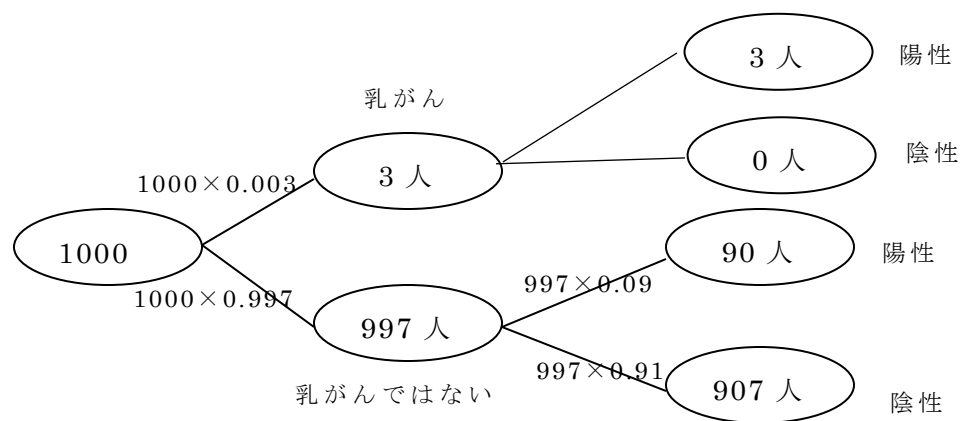
² 記入漏れは重要な問題です。記入漏れをする人は、何らかの特徴（ケアレスミスをしやすい、秘密主義者であるなど）がある人かもしれません。それらの人を記入漏れということでデータから除いてしまうと、セレクションバイアス（特定のグループが選ばれてしまうことで生じるバイアス）が発生してしまいます。

じることです³。また、面接員を前にして、道徳的また倫理的問題のある質問に正直に答えにくいことが考えられます。これ以外にも、時間面と費用面のコストもばかにできません。

8. インターネット調査の利点は、コストが低いこと、大規模調査ができること、対象者(年齢、性別など)を絞った調査ができること、回答結果が電子化されることがあります。欠点は、調査に協力してくれる人は必ずしも母集団を代表していない可能性があることです。また、同一人物による複数回答の恐れもあるでしょう。
9. 正直に答えることで人種差別主義者だと思われることを恐れて、白人有権者が黒人候補者を支持していると嘘をついた可能性があります。世論調査では、全員が正直に答えてくれるわけではありません。他人に知られたくない内容（モラル、犯罪に関わること）であれば、嘘をつく可能性が高くなります。
10. 1)1000 人中 200 人なので、20%が喫煙したことがあるといえます。2)1000 人中 500 人は表が出たから「はい」と答えています。意味がある回答は 500 人だけで、意味がある「はい」は $700-500=200$ 人だけです。よって、喫煙したことがある割合は $200/500$ で 40%と推定されます。3) 1)の調査では、回答者は嘘をついている可能性があります。2)の調査では、回答をランダム化しており、回答者の嘘をつく動機は軽減されており、より信頼のできる調査といえます⁴。
11. 1000 人のうち、乳ガンにかかっているのは 3 人（0.3%）で、全員が陽性と診断されます（下図参照）。乳ガンではない人は残り 997 人で、このうち陽性と診断されるのは 997×0.09 ですから約 90 人です。つまり、1000 人中陽性と診断される人は、 $3+90=93$ 人です。以上の結果から、陽性であり本当に乳ガンである確率は 3.2% （ $=3/93$ ）となります。

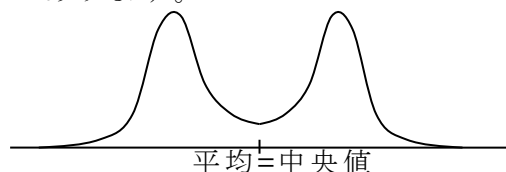
³ たとえば、面接員が調査票に適当に記入して面接したことにしてしまうかもしれません。

⁴ 回答をランダム化しても嘘をつく人はいるでしょう。しかし、回答をランダム化することで、嘘をつく動機は軽減されており、正直に話す人が増えているはずで、この調査は完璧なものではありませんが、より信頼できる調査といえます。



第 2 章の答え

1. 総和は 10140 から、平均は $10140/10=1014$ です。最頻値 1050、中央値 $(1000+1020)/2=1010$ です。偏差の 2 乗和は 143240 から、それを $9(=n-1)$ で割ると標本分散 15915.56 となり、標本標準偏差は 126.16 です。
2. 総和 $\sum_{i=1}^n x_i$ は、サンプルサイズ $n \times$ 平均 $\bar{x}$ です。この場合、合計点は 5×70 点 = 350 点です。
3. 10 人の平均は 172cm ですから、身長 of 総和は 172×10 です。総和から 160cm を除くと $172 \times 10 - 160$ です。9 人の平均は $\bar{x} = (172 \times 10 - 160)/9 = 173.33$ です。
4. 総和は表が出た回数、平均は表が出た割合と解釈できます。
5. 加重平均は $(1/3) \times 40 + (2/3) \times 90 = 73.33$ です。
6. 以下の図をみてください。明らかに左右対称の分布をしており、平均と中央値はちょうど真ん中となります。しかし、最も頻度の高いポイントである最頻値は 2 つあります。



7. ばらつきが全くない分布ですから、データは全て同じ値をとっています。
8. 1) 全ての点数に 15 点ずつ足されるため、平均は 15 点増加して $30+15=45$ となる。しかし、分布は全体的に 15 点だけ右にシフトするだけなので、標本標準偏差は変わらない。2) 全ての点数を 2 倍すると、平均も 2 倍、ばらつきも 2 倍となる。平均は $30 \times 2 = 60$ 、標本標準偏差は $5 \times 2 = 10$ となる。3) さらに 10 点追加すると、平均は 10 点増加するため $30 \times 2 + 10 = 70$ となる。しかし、分布は全体的に 10 点だけ右にシフトするので、標本標準偏差は $5 \times 2 = 10$ のままである。
9. 平均気温 95°F 、標本標準偏差 18°F から、摂氏では、平均と標本標準偏差はそれぞれ、

$$\bar{y} = -\frac{160}{9} + \left(\frac{5}{9}\right)95 = 35, \quad s_y = \left(\frac{5}{9}\right)18 = 10$$

となる。摂氏と華氏の定義は、教科書 p43 を参照してください。

10. 三郎君の英語の偏差値は $50+10\{(75-70)/5\}=60$ 、数学の偏差値は $50+10\{(66-50)/8\}=70$ です。よって、数学の方が良い成績といえます。偏差値の定義は、p46 を参照してください。

11. ここで a と b は $60=a+b50$ 、 $20=b10$ を満たすように選ばれます。よって、 $b=2$ 、 $a=-40$ となります。たとえば、日本史の点数が 60 点なら、新しい日本史の点数は $-40+2\times 60=80$ 点です。

12. a) ここで、 $\sum_{i=1}^n(x_i-\bar{x})^2$ を展開すると次のようになります。

$$\begin{aligned}\sum_{i=1}^n(x_i-\bar{x})^2 &= \sum_{i=1}^n(x_i^2+\bar{x}^2-2\bar{x}x_i) \\ &= \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \bar{x}^2 - 2\bar{x} \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i^2 + n\bar{x}^2 - 2\bar{x}n\bar{x} = \sum_{i=1}^n x_i^2 - n\bar{x}^2\end{aligned}$$

b) ここで、 $\sum_{i=1}^n(x_i-a)^2$ を展開すると、次のようになります。

$$\begin{aligned}\sum_{i=1}^n(x_i-a)^2 &= \sum_{i=1}^n((x_i-\bar{x})+(\bar{x}-a))^2 \\ &= \sum_{i=1}^n(x_i-\bar{x})^2 + n(\bar{x}-a)^2 + 2(\bar{x}-a) \sum_{i=1}^n(x_i-\bar{x})\end{aligned}$$

偏差の和は 0 なので、右辺第 3 項は 0 となります。なお、偏差の和が 0なのは、以下のように確認できます(式展開では、 $n\bar{x}=\sum_{i=1}^n x_i$ に注意)。

$$\sum_{i=1}^n(x_i-\bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = \sum_{i=1}^n x_i - n\bar{x} = 0$$

以上から、 a に依存しているのは第 2 項だけであり、 a がちょうど $\bar{x}$ のとき $\sum_{i=1}^n(x_i-a)^2$ は最小となります。

別解法として、 $\sum_{i=1}^n(x_i-a)^2$ を a で微分して 0 と置いた式を解くという方法もあります。微分に自信がない方は、本書ウェブサイトにある追加資料(微分について)を参照してください。まず、 $\sum_{i=1}^n(x_i-a)^2$ を a で微分して 0 と置きます。

$$-2 \sum_{i=1}^n(x_i-a) = 0$$

さらに、両辺を -2 で割って展開すると、次のようになります。

$$\sum_{i=1}^n x_i - n a = 0$$

これを a について解くと、 $a = \bar{x}$ が得られます。

第 3 章の答え

1. 正となる場合は、身長と体重、年齢と給料、都道府県別の人口と車保有台数、家賃と敷地面積など。負となる場合は、遊び時間と成績、家賃と築年数、家賃と駅からの距離、喫煙量と寿命など。
2. 定義式から、 x と y を入れ替えても標本相関係数は同じ。
3. 別の変数 z が、 x と y に影響を与えている「見せかけの相関」の可能性もあります。
4. x と y の両方とも平均 3 です。よって、標本共分散は、

$$s_{xy} = \frac{(1-3)(3-3) + (2-3)(2-3) + (3-3)(5-3) + (4-3)(1-3) + (5-3)(4-3)}{5-1} = 1/4$$

となり、標本標準偏差は

$$s_x = \sqrt{\frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5-1}} = \frac{\sqrt{10}}{2}$$

$$s_y = \sqrt{\frac{(3-3)^2 + (2-3)^2 + (5-3)^2 + (1-3)^2 + (4-3)^2}{5-1}} = \frac{\sqrt{10}}{2}$$

となります。以上から、標本相関係数は

$$\rho_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{1/4}{10/4} = 1/10$$

5. 標本共分散は 0 ですから標本相関係数も 0 です。式を書くと、 $\bar{x} = 0$ 、 $\bar{y} = 11.6$ から、標本共分散 s_{xy} は

$$\frac{(-5-0)(25-11.6) + (-2-0)(4-11.6) + (0-0)(0-11.6) + (2-0)(4-11.6) + (5-0)(25-11.6)}{5-1}$$

となり、これを計算すると 0 となります。標本共分散は 0 ですから、標本相関係数も 0 となります。

6. 強いチームなら人気も上がり収入も増えますから、多額の年俸を支払うことができるといえます。また、多額の年俸を支払えば良い選手を高額で引き抜くことができ、強いチームになるともいえます。よって、①チームの強さ→年俸、②年俸（選手のやる気がでたり、補強を行える）→チームの強さという双方からの因果関係がありそうです。相関関係を見ても、①と②の効果が混在しているため、②の効果が本当にあるのか、また、どの程度あるのかは分かりま

せん。

7. 日本は年功序列が一般的です。よって、年齢とともに所得は上昇する傾向があります。年齢とともに血圧は上昇する傾向があります。以上から、両者の関係は「見せかけの相関」と考えられます。

8. 間違っています。アイスクリームの消費も夏に増える傾向があり、「見せかけの相関」の可能性もあります。

9. 1)学科別の合格率をみると、男女によって合格率の差はほとんどないか、もしくは女性の方が高くなっています。全体の合格率に差が出た理由は、男性が難易度の低い学科を志望し、逆に、女性が難易度の高い学科を志望しているからです。たとえば、男性志願者は難易度の低い学科 A、B に多く、女性志願者は難易度の高い学科 C、E に多くなっています。この例のように、全体の傾向と層別の傾向が異なってしまう現象は、**シンプソンのパラドックス (Simpson's paradox)** といわれます。

2)男性の合格率は、男性だけの志願者数を加重としています。これは以下のよう
に確認できます。

$$\begin{aligned}\frac{512+353+120+138+53+22}{2691} &= \frac{825}{2691} \frac{512}{825} + \frac{560}{2691} \frac{353}{560} + \frac{325}{2691} \frac{120}{325} + \frac{417}{2691} \frac{138}{417} + \frac{191}{2691} \frac{53}{191} + \frac{373}{2691} \frac{22}{373} \\ &= \frac{825}{2691} \cdot 0.621 + \frac{560}{2691} \cdot 0.630 + \frac{325}{2691} \cdot 0.369 + \frac{417}{2691} \cdot 0.331 + \frac{191}{2691} \cdot 0.277 + \frac{373}{2691} \cdot 0.059 = 0.445\end{aligned}$$

同様に、女性の合格率は女性の志願者数を加重としています。

$$\frac{108}{1835} \cdot 0.824 + \frac{25}{1835} \cdot 0.680 + \frac{593}{1835} \cdot 0.341 + \frac{375}{1835} \cdot 0.349 + \frac{393}{1835} \cdot 0.239 + \frac{341}{1835} \cdot 0.070 = 0.304$$

3)男女全体の受験者数で加重を付けると

$$\begin{aligned}\frac{933}{4526} \cdot 0.621 + \frac{585}{4526} \cdot 0.630 + \frac{918}{4526} \cdot 0.369 + \frac{792}{4526} \cdot 0.331 + \frac{584}{4526} \cdot 0.277 + \frac{714}{4526} \cdot 0.059 &= 0.387 \\ \frac{933}{4526} \cdot 0.824 + \frac{585}{4526} \cdot 0.680 + \frac{918}{4526} \cdot 0.341 + \frac{792}{4526} \cdot 0.349 + \frac{584}{4526} \cdot 0.239 + \frac{714}{4526} \cdot 0.070 &= 0.430\end{aligned}$$

であり、男女別の合格率にほとんど差はありません。

10. 2001 年 9 月に発症した BSE によって、2002 年に牛肉の輸入量が急減しました。BSE 問題は 2003 年には沈静化し、牛肉輸入量はもとの水準に回復しまし

た。BSE が原因で 2003 年の輸入量の増加率は一時的に高くなっており、セーフガードの発動は不適切だったと考えられます。

11. (1) 偏差の積和は以下として展開できます。

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n (x_i y_i + \bar{x}\bar{y} - \bar{x}y_i - \bar{y}x_i) = \sum_{i=1}^n x_i y_i + n\bar{x}\bar{y} - \bar{x} \sum_{i=1}^n y_i - \bar{y} \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i y_i + n\bar{x}\bar{y} - \bar{x}n\bar{y} - \bar{y}n\bar{x} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}\end{aligned}$$

(2) 偏差の積和は以下として表現できます。

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i - \sum_{i=1}^n (x_i - \bar{x})\bar{y}$$

偏差の和は 0 ですから、上式 2 項目は 0 となります。つまり、

$$\sum_{i=1}^n (x_i - \bar{x})\bar{y} = \bar{y} \sum_{i=1}^n (x_i - \bar{x}) = \bar{y} \times 0 = 0$$

です。よって、 $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i$ となります。同様に、

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (y_i - \bar{y})x_i - \sum_{i=1}^n (y_i - \bar{y})\bar{x}$$

と書けます。偏差の和は 0 から、上式 2 項目は 0 となります。よって、

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (y_i - \bar{y})x_i$$

第 4 章の答え

1. この計算は同時確率を考慮していないため誤っています。
2. A 大学に合格する事象を A とし、B 大学に合格する事象を B とします。それぞれの確率は $P\{A\}=0.1$, $P\{B\}=0.05$, $P\{A \cap B\}=0.03$ ですから、少なくともどちらかの大学に合格 ($A \cup B$) する確率は、 $P\{A \cup B\}=P\{A\}+P\{B\}-P\{A \cap B\}=0.1+0.05-0.03=0.12$ です。
3. 加法定理から $P\{A_1 \cup A_2\}=P\{A_1\}+P\{A_2\}-P\{A_1 \cap A_2\}=2 \times (1/6)-(1/6)^2=11/36$ です。ド・メレの間違ひは同時確率 (2 回とも 6 の目が出る確率) を引かなかったことです。
4. ここで 1 個目の玉が赤である事象を A、2 番目の玉が赤である事象を B とします。両方とも赤は、 $P\{A \cap B\}=P\{A\}P\{B|A\}$ です。玉は合計 9 個あり、そのうち赤は 4 個ですから $P\{A\}=4/9$ です。1 個目が赤なら、残り 8 個のうち 3 個が赤ですから $P\{B|A\}=3/8$ です。よって、 $P\{A \cap B\}=(4/9) \times (3/8)=1/6$ 。
5. 事象 A を試合に勝つ、事象 B を雨が降るとします。このとき、 $P\{B\}=0.3$ 、 $P\{\bar{B}\}=0.7$ 、 $P\{A|B\}=0.7$ 、 $P\{A|\bar{B}\}=0.9$ です。試合に勝つ確率は、
 $P\{A\}=P\{A \cap B\}+P\{A \cap \bar{B}\}=P\{A|B\}P\{B\}+P\{A|\bar{B}\}P\{\bar{B}\}=0.7 \times 0.3+0.9 \times 0.7=0.84$ 。
6. 52 枚から 5 枚を取り出すとき組み合わせは ${}_{52}C_5$ 通りあります (付録 A 参照)。取り出された 5 枚が全てハートであるとき、その組み合わせは計 ${}_{13}C_5$ 通りあります。したがって、52 枚から 5 枚取り出して全てハートの確率は、

$$\begin{aligned} \frac{{}_{13}C_5}{{}_{52}C_5} &= \left(\frac{13!}{5!(13-5)!} \right) / \left(\frac{52!}{5!(52-5)!} \right) \\ &= \frac{13 \times 12 \times 11 \times 10 \times 9}{52 \times 51 \times 50 \times 49 \times 48} = 0.000495 \end{aligned}$$

7. 1) A_1 を全員が男、 A_2 を全員が女と定義します。ここで $P\{A_1\}$ は 3 人全てが男の子である確率です。3 人全てが男である確率は $(1/2)^3$ です。同様に、3 人全てが女である確率は $(1/2)^3$ です。 A_1 と A_2 が互いに排反ですから、全部の子供が同姓の確率は、 $P\{A_1 \cup A_2\}=P\{A_1\}+P\{A_2\}=(1/2)^3+(1/2)^3=(1/2)^2$ 。2) A_i を i 番目が女で他が男の事象と新たに定義します (たとえば、 A_1 は女男

男)。A₁、A₂、A₃は互いに排反ですから、女1人、男2人となる確率は
 $P\{A_1 \cup A_2 \cup A_3\} = P\{A_1\} + P\{A_2\} + P\{A_3\} = (1/2)^3 + (1/2)^3 + (1/2)^3 = 3/8$ 。

8. HIV感染者をA₁、HIV非感染者をA₂、陽性反応をBとします。それぞれの確率は、 $P\{A_1\}=0.001$ 、 $P\{A_2\}=0.999$ 、 $P\{B|A_1\}=1.0$ 、 $P\{B|A_2\}=0.01$ です。このとき、陽性反応の確率は
 $P\{B\} = P\{B|A_1\}P\{A_1\} + P\{B|A_2\}P\{A_2\}$
 $= 1.0 \times 0.001 + 0.01 \times 0.999 = 0.01099$ 。よって、
 $(P\{B|A_1\}/P\{B\})P\{A_1\} = (1.0/0.01099) \times 0.001 = 0.0909$

9. 同時確率は以下として書けます。

$$P\{A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap A_n\} = P\{A_1\} \frac{P\{A_1 \cap A_2\}}{P\{A_1\}} \dots \frac{P\{A_1 \cap A_2 \cap \dots \cap A_{n-1}\}}{P\{A_1 \cap A_2 \cap \dots \cap A_{n-2}\}} \frac{P\{A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap A_n\}}{P\{A_1 \cap A_2 \cap \dots \cap A_{n-1}\}}$$

両辺が等しくなるのは、右辺の分母と分子が上手く打ち消し合っているからです。また、条件付き確率の定義から以下が成立します。

$$P\{A_2 | A_1\} = \frac{P\{A_1 \cap A_2\}}{P\{A_1\}}, \dots,$$

$$P\{A_{n-1} | A_1 \cap A_2 \cap \dots \cap A_{n-2}\} = \frac{P\{A_1 \cap A_2 \cap \dots \cap A_{n-2} \cap A_{n-1}\}}{P\{A_1 \cap A_2 \cap \dots \cap A_{n-2}\}}$$

$$P\{A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}\} = \frac{P\{A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap A_n\}}{P\{A_1 \cap A_2 \cap \dots \cap A_{n-1}\}}$$

これらを同時確率の式に代入すると、同時確率が条件付き確率の積として表現できることが分かります。

$$P\{A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap A_n\} = P\{A_1\}P\{A_2|A_1\} \dots P\{A_{n-1}|A_1 \cap \dots \cap A_{n-2}\}P\{A_n|A_1 \cap \dots \cap A_{n-1}\}$$

10. 囚人1が恩赦ならA₁、囚人2が恩赦ならA₂、囚人3が恩赦ならA₃、「囚人2が処刑される」といわれる場合をBとします。それぞれの確率は、 $P\{A_1\}=P\{A_2\}=P\{A_3\}=1/3$ です。囚人1が恩赦なら、看守は囚人2も3も処刑のためランダムに囚人2が処刑といった、と考えられます

($P\{B|A_1\}=0.5$)。囚人2が恩赦なら、囚人2が処刑といわれることはありません ($P\{B|A_2\}=0.0$)。囚人3が恩赦なら、囚人2が処刑といわれます

($P\{B|A_3\}=1.0$)。以上から、

$$P\{B\} = P\{B|A_1\}P\{A_1\} + P\{B|A_2\}P\{A_2\} + P\{B|A_3\}P\{A_3\} = 0.5 \times (1/3) + 0.0 \times (1/3) + 1.0 \times (1/3) = 0.5$$

よって、 $P\{A_1|B\} = (P\{B|A_1\}/P\{B\})P\{A_1\} = (0.5/0.5) \times (1/3) = 1/3$ 。

したがって、看守は「囚人 2 が処刑される」といったとき、囚人 1 が恩赦となる確率は $1/3$ から変わっていません。また、囚人 3 に関しては、 $P\{A_3|B\}=(P\{B|A_3\}/P\{B\})P\{A_3\}=(1.0/0.5)\times(1/3)=2/3$ から、看守は「囚人 2 が処刑される」といったとき、囚人 3 が恩赦となる確率は $1/3$ から $2/3$ に上がっています。

11. 1 番に車が入っている場合を事象 A_1 、2 番に車が入っている場合を事象 A_2 、3 番に車が入っている場合を事象 A_3 とします。そして、「司会者が 2 番のドアを開ける」という場合を事象 B とします。それぞれの確率は $P\{A_1\}=P\{A_2\}=P\{A_3\}=1/3$ です。もし 1 番のドアに車が入っているなら、司会者は 2、3 番のどちらを開けてもよいので、ランダムにどちらかを開けます ($P\{B|A_1\}=0.5$)。もし 2 番のドアに車が入っているなら、2 番のドアを開けるとクイズが終わるため、司会者は 3 番のドアを開けます ($P\{B|A_2\}=0.0$)。もし 3 番のドアに車が入っているなら、司会者は 2 番のドアを開けます ($P\{B|A_3\}=1.0$)。以上から、司会者が 2 番のドアを開ける確率は、 $P\{B\}=P\{B|A_1\}P\{A_1\}+P\{B|A_2\}P\{A_2\}+P\{B|A_3\}P\{A_3\}=0.5\times(1/3)+0.0\times(1/3)+1.0\times(1/3)=0.5$ です。よって、2 番のドアが開けられた後で 1 番のドアに車が入っている確率は、 $P\{A_1|B\}=(P\{B|A_1\}/P\{B\})P\{A_1\}=(0.5/0.5)\times(1/3)=1/3$ で変わりません。また、2 番のドアが開けられた後で 3 番のドアに車が入っている確率は、 $P\{A_3|B\}=(P\{B|A_3\}/P\{B\})P\{A_3\}=(1.0/0.5)\times(1/3)=2/3$ です。以上から、1 番のドアであれば当選確率は $1/3$ で、ドアを変えると $2/3$ に増えるのです。

12. たとえば、左前輪といって、これが偶然に当たる確率は $1/16$ です。タイヤは計 4 つですから、偶然当たる確率は $4\times(1/16)=1/4$ です。

13. 確率の公理とは、

- (1) 任意の事象 A に対して $0\leq P\{A\}$
- (2) Ω に対して $P\{\Omega\}=1$
- (3) A と B が互いに排反ならば $P\{A\cup B\}=P\{A\}+P\{B\}$

となります。以下では、先験的確率、経験的確率、条件付き確率について確率の公理が全て満たされることを証明します。

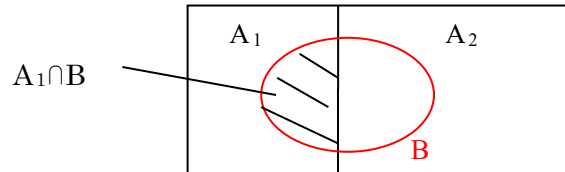
まずは、先験的確率が確率の公理を満たすことを確認しましょう。公理(1)

については、 $n > 0$ と $n(A) \geq 0$ ですから $P\{A\} = n(A)/n \geq 0$ となります。公理(2)は、 $n(\Omega) = n$ ですから、 $P\{\Omega\} = n(\Omega)/n = 1$ です。公理(3)についても、 A と B が互いに排反であれば、 A と B に何らの共通点もありませんから、 $n(A \cup B) = n(A) + n(B)$ です。よって、 $P\{A \cup B\} = n(A \cup B)/n = n(A)/n + n(B)/n = P\{A\} + P\{B\}$ となります。

経験的確率も、確率の公理を満たします。公理(1)については、 $n > 0$ 、 $n(A) \geq 0$ ですから、 $P\{A\} = n(A)/n \geq 0$ が成立します。次に、公理(2)について考えましょう。何度試行をしても、その結果は標本空間 Ω のどれかが生じます。よって、観察回数 n は $n(\Omega)$ に等しくなります。よって、 $P\{\Omega\} = n(\Omega)/n = 1$ です。公理(3)についても、 A と B が互いに排反であれば、 $n(A \cup B) = n(A) + n(B)$ です。たとえば、コインを投げて、表か裏が出た回数を求めるには、表が出た回数+裏が出た回数を求めればよいです。よって、

$$P\{A \cup B\} = n(A \cup B)/n = n(A)/n + n(B)/n = P\{A\} + P\{B\}$$

条件付き確率も確率の公理を満たします。公理(1)については、 $P\{A \cap B\} \geq 0$ 、 $P\{B\} > 0$ ですから、 $P\{A|B\} = P\{A \cap B\}/P\{B\} \geq 0$ です。また、公理(2)については、 $P\{B \cap \Omega\} = P\{B\}$ ですから、 $P\{\Omega|B\} = P\{B \cap \Omega\}/P\{B\} = P\{B\}/P\{B\} = 1$ となります。公理(3)について考えましょう。2つの排反する事象 A_1 と A_2 を考えます。下図の通り、事象 A_1 と A_2 は排反しているため、標本空間は A_1 と A_2 に分けられます。



新たに事象 B を考えると、 $(A_1 \cup A_2) \cap B = (A_1 \cap B) \cup (A_2 \cap B)$ となります。 A_1 と A_2 が互いに排反ですから、 $(A_1 \cap B)$ と $(A_2 \cap B)$ も互いに排反です。よって、

$$\begin{aligned} P\{A_1 \cup A_2|B\} &= P\{(A_1 \cup A_2) \cap B\} / P\{B\} \\ &= P\{(A_1 \cap B) \cup (A_2 \cap B)\} / P\{B\} \\ &= (P\{A_1 \cap B\} + P\{A_2 \cap B\}) / P\{B\} \\ &= P\{A_1 \cap B\} / P\{B\} + P\{A_2 \cap B\} / P\{B\} \\ &= P\{A_1|B\} + P\{A_2|B\} \end{aligned}$$

14. ここでは独立と排反が異なる概念であることを2つの方法で示します。

排反と独立とは異なる概念ですが、混同しやすい概念ですから気を付けて

ください。

- 1) 事象 A と B が独立なら、 $P\{A \cap B\} = P\{A\}P\{B\}$ が成立します。ここで $P\{A\} > 0$ 、 $P\{B\} > 0$ なら、積事象の確率はプラス ($P\{A \cap B\} = P\{A\}P\{B\} > 0$) になります。これに対し、事象 A と B が排反なら、両事象に何らの共通点也没有から ($A \cap B$ は空集合 ϕ)、積事象の確率は 0 になります ($P\{A \cap B\} = P\{\phi\} = 0$)。つまり、A と B が排反なら $P\{A \cap B\} = P\{\phi\} = 0 \neq P\{A\}P\{B\}$ となり、両者は独立になりません。
- 2) 事象 A と B が独立なら、 $P\{A\} = P\{A|B\}$ が成立します。つまり、事象 B の情報は事象 A の確率に影響を与えません。これに対し、事象 A と B が排反なら、事象 B の情報は事象 A の確率に影響を与えます。事象 A と B が排反なら $P\{A \cap B\} = 0$ ですから、事象 A の条件付き確率は 0 になります ($P\{A|B\} = P\{A \cap B\}/P\{B\} = 0$)。ここで $P\{A\} > 0$ なら、 $P\{A\} \neq P\{A|B\}$ となります。

第 5 章の答え

- 1測定単位が 1 円なら、10000X から、期待値は $10000 \times 0.25 = 2500$ 、分散は $10000^2 \times 0.5625 = 56250000$ 、標準偏差は $10000 \times 0.75 = 7500$ 。どちらの測定単位で答えても正解。スケールの問題が無い期待値と標準偏差はどちらも同じ値となっている。ただし、スケールの問題がある分散は、単位として何を用いるかで結果が違う。
- 2測定単位が 1 円なら、期待値 71000、分散 189000000、標準偏差 13748。どちらの測定単位で答えても正解。

170.88。X と Y に強い正の相関があり、夫婦の年収の標準偏差も高くなっています。

- 8 X+Y と X-Y の共分散は、

$$\begin{aligned}
 \text{Cov}(X+Y, X-Y) &= E[(X+Y-E[X+Y])(X-Y-E[X-Y])] \\
 &= E[\{(X-E[X])+(Y-E[Y])\}\{(X-E[X])-(Y-E[Y])\}] \\
 &= E[(X-E[X])^2] - E[(Y-E[Y])^2] + E[(X-E[X])(Y-E[Y])] - E[(X-E[X])(Y-E[Y])] \\
 &= E[(X-E[X])^2] - E[(Y-E[Y])^2] \\
 &= V(X) - V(Y)
 \end{aligned}$$

となります。ここで $V(X)=V(Y)$ なら $\text{Cov}(X+Y, X-Y)=0$ となります。

- 9 練習問題 8 において、 $X^*=X+Y$ 、 $Y^*=X-Y$ と定義すると、 X^* と Y^* はともに X と Y の関数ですから独立ではありませんが、両者の共分散は 0 となります。

別の例として、確率変数 Z_1 と Z_2 の期待値はともに 0 ($E[Z_1]=E[Z_2]=0$)、互いに独立としましょう ($E[Z_1Z_2]=0$)。ここで $X=Z_1Z_2$ 、 $Y=Z_2$ と定義すると、X と Y はともに Z_2 の関数ですから両者は独立ではありません。しかし、 $E[X]=E[Z_1]E[Z_2]=0$ 、 $E[Y]=E[Z_2]=0$ から、共分散は 0 となります。

$$\begin{aligned}
 \text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\
 &= E[XY] - 0 \\
 &= E[Z_1Z_2^2] \\
 &= E[Z_1]E[Z_2^2] = 0 \times E[Z_2^2] = 0
 \end{aligned}$$

式展開で、 Z_1 と Z_2 は独立ですから、 Z_1 と Z_2^2 も独立であることを用いました。

独立なら共分散は 0 となりますが、共分散が 0 でも独立を意味しないことを覚えておいてください。独立は共分散よりも、両変数が無関係であることを示す強い概念となります。

- 10 1) 期待収益率は、 $E[\omega r_s + (1-\omega) r_b] = \omega E[r_s] + (1-\omega)E[r_b] = \omega 0.08 + (1-\omega)0.03$ です。よって、全てを株式で運用することで ($\omega=1$)、収益率は最大の 8% となります。

2) $\omega r_s + (1-\omega)r_b$ の分散を考えましょう。株式の標準偏差は 0.07、債券の標準偏差 0.03、相関係数は 0.05 ですから、相関係数の定義から共分散

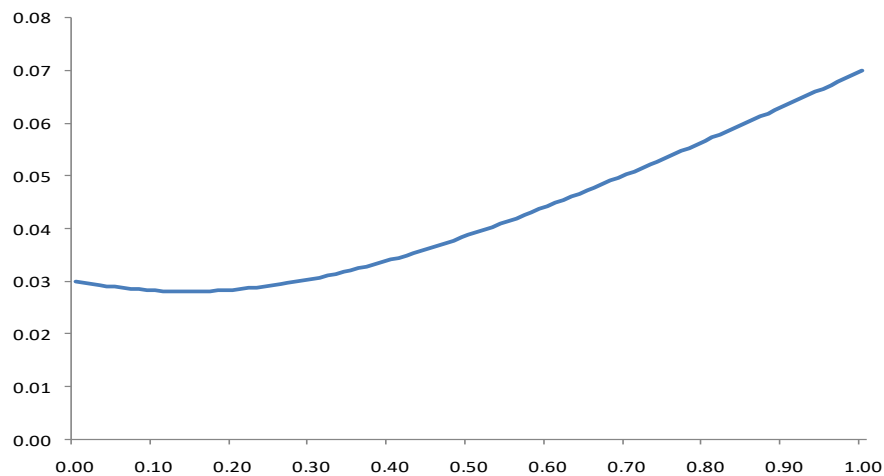
$\text{Cov}(r_s, r_b) = \rho_{rs,rb} \sqrt{V(r_s)V(r_b)} = 0.05 \times 0.07 \times 0.03 = 0.000105$ です。この資産の分散は、

$$\begin{aligned} V(\omega r_s + (1-\omega)r_b) &= E[\{(\omega r_s + (1-\omega)r_b) - (\omega 0.08 + (1-\omega)0.03)\}^2] \\ &= E[\{(\omega(r_s - 0.08) + (1-\omega)(r_b - 0.03))\}^2] \\ &= E[\omega^2(r_s - 0.08)^2 + (1-\omega)^2(r_b - 0.03)^2 + 2\omega(1-\omega)(r_s - 0.08)(r_b - 0.03)] \\ &= \omega^2 E[(r_s - 0.08)^2] + (1-\omega)^2 E[(r_b - 0.03)^2] + 2\omega(1-\omega)E[(r_s - 0.08)(r_b - 0.03)] \end{aligned}$$

ここで $E[(r_s - 0.08)^2] = V(r_s) = 0.07^2$ 、 $E[(r_b - 0.03)^2] = V(r_b) = 0.03^2$ 、 $E[(r_s - 0.08)(r_b - 0.03)] = 0.000105$ となるため、上式に、これらの値を代入することで

$$\begin{aligned} &\omega^2 0.07^2 + (1-\omega)^2 0.03^2 + 2\omega(1-\omega)0.000105 \\ &= 0.0049\omega^2 + (0.0009 + 0.0009\omega^2 - 0.0018\omega) + (0.00021\omega - 0.00021\omega^2) \\ &= 0.00559\omega^2 - 0.00159\omega + 0.0009 \end{aligned}$$

下図では、上式の平方根を取った標準偏差を縦軸、横軸を ω として示しています。 ω が 0.14 で標準偏差が 0.028 となり最小となっています。つまり、株で資産の 14%、債権で 86% を運用することで、標準偏差を最小化できます³。このとき期待収益率は、 $0.14 \times 0.08 + (1-0.14)0.03 = 0.037$ です。もっとも安全な資産運用は、すべてを債券だけで運用することではありません（債権だけだと、期待収益率 3%、標準偏差 3%）。



11

1) 確率変数 X の期待値は μ 、分散は σ^2 でり、実現値は $x_1, x_2, \dots, x_m$ としましよ

³ 分散の式を ω で微分して 0 と置くと、 $2 \times 0.00559\omega - 0.00159 = 0$ ですから $\omega = 0.14$ が得られます。

う。確率変数 $Z = X - \mu$ と定義すると、 Z の実現値は $z_1 = x_1 - \mu, \dots, z_m = x_m - \mu$ となります。 Z には m 個の実現値がありますが、その中で $|z_i| \geq \kappa\sigma$ を満たす z_i だけからなる集合を A とします。

ここで

$$P\{|Z| \geq \kappa\sigma\} = \sum_{z \in A} P\{z_i\}$$

になります。 Σ 記号の $z \in A$ は、集合 A の中にある z_i について、確率 $P\{z_i\}$ の和をとることを意味します。集合 A の中にある z_i は $|z_i| \geq \kappa\sigma$ を満たすため、両辺を 2 乗した $z_i^2 \geq \kappa^2\sigma^2$ も満たします（つまり、 $z_i^2/(\kappa^2\sigma^2) \geq 1$ ）。このため、上式右辺は

$$\sum_{z \in A} P\{z_i\} \leq \sum_{z \in A} \frac{z_i^2}{\kappa^2\sigma^2} P\{z_i\} \leq \sum_{i=1}^m \frac{z_i^2}{\kappa^2\sigma^2} P\{z_i\} = \frac{1}{\kappa^2\sigma^2} \sum_{i=1}^m z_i^2 P\{z_i\}$$

となります。2 番目の不等式は、集合 A だけでなく、全ての z_i について和をとっているため、大きくなっています。最後の等号は定数 $1/(\kappa^2\sigma^2)$ を外に出しています。

以上から、

$$P\{|Z| \geq \kappa\sigma\} \leq \frac{1}{\kappa^2\sigma^2} \sum_{i=1}^m z_i^2 P\{z_i\}$$

となります。ここで左辺は $P\{|Z| \geq \kappa\sigma\} = P\{|X - \mu| \geq \kappa\sigma\}$ であり、右辺は

$$\sum_{i=1}^m z_i^2 P\{z_i\} = \sum_{i=1}^m (x_i - \mu)^2 P\{x_i\} = \sigma^2$$

になるため（ $z_i = x_i - \mu$ 、 X の分散は σ^2 に注意）、

$$P\{|X - \mu| \geq \kappa\sigma\} \leq \frac{1}{\kappa^2\sigma^2} \sigma^2 = \frac{1}{\kappa^2}$$

になります。これで証明は終わりです。

2) 確率の和は 1 から、

$$P\{|X - \mu| < \kappa\sigma\} + P\{|X - \mu| \geq \kappa\sigma\} = 1$$

であり、 $P\{|X - \mu| \geq \kappa\sigma\}$ を右辺に移項させると、

$$P\{|X - \mu| < \kappa\sigma\} = 1 - P\{|X - \mu| \geq \kappa\sigma\}$$

となります。チェビシェフの不等式 $P\{|X - \mu| \geq \kappa\sigma\} \leq 1/\kappa^2$ を用いると、この式は

$$P\{|X - \mu| < \kappa\sigma\} = 1 - P\{|X - \mu| \geq \kappa\sigma\}$$

$$\geq 1 - \frac{1}{\kappa^2}$$

となります。ここで上式の左辺は

$$\begin{aligned} P\{|X - \mu| < \kappa\sigma\} &= P\{-\kappa\sigma < X - \mu < \kappa\sigma\} \\ &= P\{\mu - \kappa\sigma < X < \mu + \kappa\sigma\} \end{aligned}$$

と書けます。これまでの結果から、

$$P\{\mu - \kappa\sigma < X < \mu + \kappa\sigma\} \geq 1 - \frac{1}{\kappa^2}$$

が得られました。

3) 2)の結果から、「確率変数 X が $\mu \pm \kappa\sigma$ の領域に留まる割合は $1 - 1/\kappa^2$ 以上になる」といえます。この関係を使えば、確率変数 X がどのような分布に従っていても、期待値 μ から標準偏差 σ の $\pm\kappa$ 倍以内に留まる割合が推測できることになるわけです。たとえば、 $\kappa = 1$ なら

$$P\{\mu - \sigma < X < \mu + \sigma\} \geq 1 - \frac{1}{1^2} = 0$$

ですから、0%以上の確率で、 X は期待値 μ から $\pm\sigma$ 以内に留まる、ことがわかります（これは当たり前のことですが・・・）。また、 $\kappa = 2$ なら、

$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} \geq 1 - \frac{1}{2^2} = \frac{3}{4}$$

ですから、75%(=3/4)以上の確率で、 X は期待値 μ から $\pm 2\sigma$ 以内に留まります。同様に、 $\kappa = 3$ なら

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} \geq 1 - \frac{1}{3^2} = \frac{8}{9}$$

ですから、89%(=8/9)以上の確率で、 X は期待値 μ から $\pm 3\sigma$ 以内に留まります。

教科書の第2章(p45)では、分布が釣鐘状のとき、平均と標本標準偏差を使って範囲と割合の関係を紹介しました。しかし、この関係は全ての分布に成立する訳ではありません。これに対し、チェビシェフの不等式は、確率変数に期待値と分散が存在しているなら、どのような分布でも成立する一般的関係式とな

ります。

当然ですが、釣鐘状の分布に限定したときの方が、範囲と割合について、平均と標準偏差はより多くの情報を持っています。たとえば、 $\mu \pm 2\sigma$ であれば、釣鐘状分布では約 95%がその中に含まれると言えますが、チェビシェフの不等式では 75%以上が含まれるとしか言えません。チェビシェフの不等式は、前提条件が一般的であるがゆえに、弱い含意しか持っていないのです。残念ながら、この世にフリーランチは存在しません (There's no such thing as a free lunch)。前提条件が弱いなら、その含意は弱くなるのは当たりまえのことでしょう。

第 6 章の答え

- 1 $E[X]=1p+0(1-p)=p$ 、 $E[X^2]=1^2p+0^2(1-p)=p$ となりますから、分散は $V(X)=p-p^2=p(1-p)$ となります。
- 2 1) X は品質が良ければ 1 となるベルヌーイ確率変数とします ($p=0.7$) 中古車の価値は $80X+30(1-X)$ ですから、その期待値は $E[80X+30(1-X)]=80 \times 0.7+30 \times 0.3=65$ 万円 (市場価格)。2) 売り手は品質の良い車に 80 万円の価値があることを知っており、価格 65 万円では悪い品質の車だけを販売します。よって、中古車市場では悪い品質の車だけが取引され、良い品質の車は市場から無くなってしまいます (経済学では、このような市場をレモン市場といいます)。この問題を解決するためには、情報の非対称性を取り除く、つまり、売り手が買い手に中古車の正確な情報を開示する必要があります。
- 3 1) $X \sim B(10, 1/3)$ なら $E[X]=10/3$ です。 $E[10X-5(10-X)]=10E[X]-5(10-E[X])=10 \times 10/3-5(10-10/3)=0$ から適当に答えても平均 0 です。2) 10 問中 x 問正解する確率は、

$$P\{X = x\} = \frac{10!}{x!(10-x)!} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{10-x}$$

です。総得点が正とは $(10x-5(10-x)>0)$ 、 $x>50/15$ のときですから、4 問以上に正しく回答すれば総得点は正です。

$$P\{X \geq 4\} = 1 - \sum_{i=0}^3 P\{X = i\}$$

これを計算すると、 $P\{X \geq 4\}=0.44$ です。

- 4 二項確率変数は相互に独立なベルヌーイ確率変数 (1 となる確率は p 、0 となる確率は $1-p$) の和となります。しかし、相撲における試合結果は互いに独立ではないでしょう。調子の良し悪しがあり、前の試合の結果が今日の結果に影響するでしょう。また、相手の強さに応じて、勝率も一定ではなく、毎回変わると考えられます。以上から、勝利数 X を相互に独立なベルヌーイ確率変数の和と考えるのは厳密には正しくありません。
- 5 今日雨が降ると、明日も雨が降りやすくなるものです。また、時期によって降水確率も一定でないでしょう。こう考えると、雨の降る回数 X を相互

に独立なベルヌーイ確率変数の和と考えるのは正しくありません¹。

- 6 二項分布は相互に独立なベルヌーイ確率変数の和であることを思い出してください。ここで $X=X_1+X_2+\dots+X_n$ 、 $Y=Y_1+Y_2+\dots+Y_m$ であり、 X_i と Y_j は相互に独立なベルヌーイ確率変数（ともに p の確率で 1 となり、 $1-p$ の確率で 0 となる）とします。このとき、 $X+Y=X_1+X_2+\dots+X_n+Y_1+Y_2+\dots+Y_m$ であり、各要素は相互に独立な $n+m$ 個のベルヌーイ確率変数の和ですから、これは 2 項分布 $B(n+m,p)$ に従います²。

$$7 \quad 1) \quad P\{X > 10\} = P\left\{\frac{X-12}{2} > \frac{10-12}{2}\right\} = P\{Z > -1\}$$

また、分布表から $P\{Z > -1\} = P\{Z < 1\} = 0.8413$ となります。

$$2) \quad P\{X < 8\} = P\left\{\frac{X-12}{2} < \frac{8-12}{2}\right\} = P\{Z < -2\}$$

また、分布表から $P\{Z < -2\} = P\{Z > 2\} = 1 - P\{Z < 2\} = 1 - 0.9772 = 0.0228$ となります。

3) 左辺は 12 で引かれていますから、標準化は両辺を 2 で割ります。

$$P\{|X-12| < 3\} = P\left\{\left|\frac{X-12}{2}\right| < \frac{3}{2}\right\} = P\{|Z| < 1.5\}$$

$P\{|Z| < 1.5\} = P\{Z < 1.5\} - P\{Z < -1.5\} = P\{Z < 1.5\} - (1 - P\{Z < 1.5\}) = 0.9332 - (1 - 0.9332) = 0.8664$ となります。

$$\begin{aligned} 8 \quad P\{220 < X < 280\} &= P\left\{\frac{220-230}{20} < \frac{X-230}{20} < \frac{280-230}{20}\right\} \\ &= P\{-0.5 < Z < 2.5\} = P\{Z < 2.5\} - P\{Z < -0.5\} \\ &= P\{Z < 2.5\} - (1 - P\{Z < 0.5\}) = 0.9938 - (1 - 0.6915) = 0.6853 \end{aligned}$$

- 9 $X \sim B(14, 1/2)$ ですから、1) $X=4$ の確率は

$$P\{X = 4\} = \frac{14!}{4!(14-4)!} (1/2)^4 (1/2)^{14-4} = 0.0611$$

2) $np=14(1/2)=7$ 、 $npq=14(1/2)^2=3.5$ から、二項分布は $X \sim N(7, 3.5)$ で正規近似できます。したがって、

¹ 気候が急激に変化しているとき、過去のデータから雨の確率 p を推定するのは問題です。この場合、至近のデータだけを用いて p を推定する必要があります。

² 厳密には、分布関数が一致することで証明します。

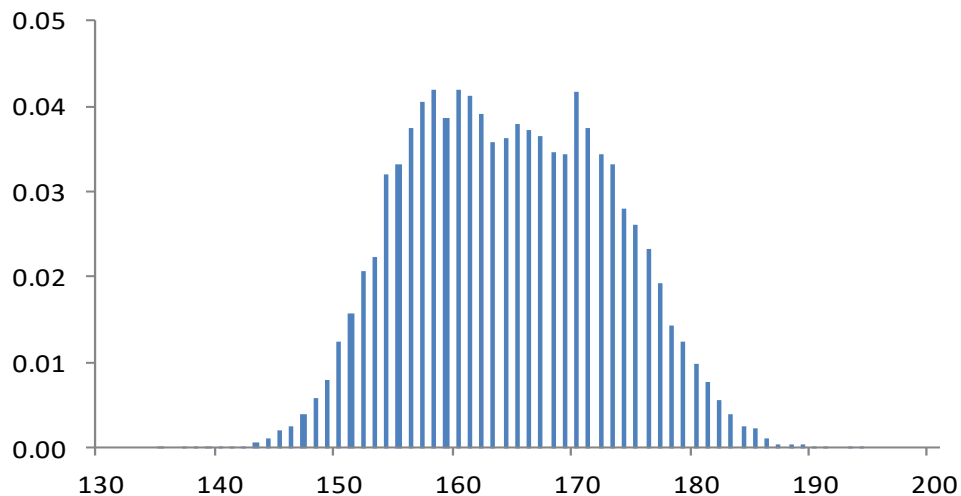
$$P\{3.5 < X < 4.5\} = P\left\{\frac{3.5-7}{\sqrt{3.5}} < \frac{X-7}{\sqrt{3.5}} < \frac{4.5-7}{\sqrt{3.5}}\right\}$$

$$= P\{-1.87 < Z < -1.34\} = 0.0594$$

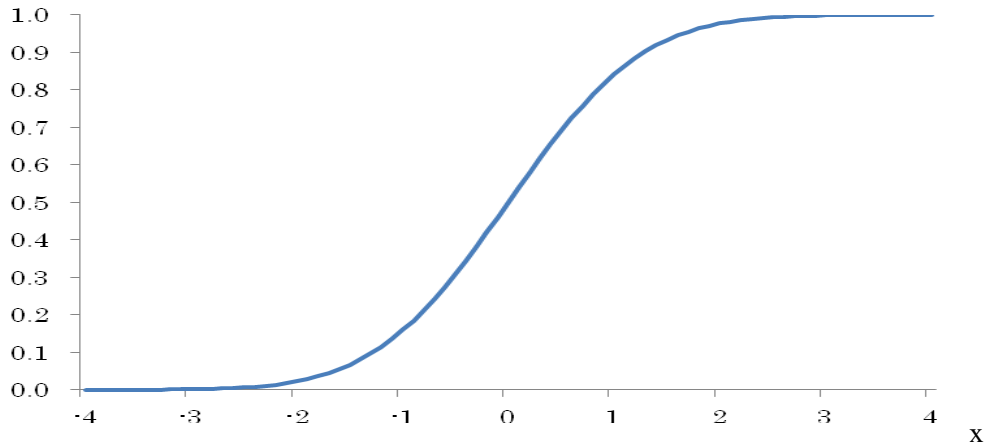
10 適切に設計された試験では、点数は正規分布に従うものです。もし 80 点で人数が極端に多ければ、何らかの不正が行われた可能性があります。この場合、80 点前後の答案を調べて、何らかの不正が行われた可能性を分析する必要があります。

11 男女全てを 1 つのグループとして身長分布を図示しました（男女の身長の和ではありません）。男女全体の身長分布は、男性と女性の身長分布を混合させた分布（混合分布）となっています。頂点が 2 つあるのは、男性平均 170cm と女性平均 158cm で頂点が 2 つ形成されているためです。

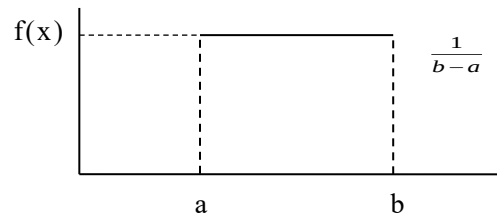
正規分布の性質から、正規確率変数の線形結合は正規分布に従いました。たとえば、男子の身長 + 女子の身長は正規分布に従います。しかし、男子と女子を 1 つのグループとするということは、正規確率変数の線形結合ではないため、男女全員の身長分布は正規分布に従いません。



12 標準正規確率変数の累積分布関数は以下の通りです。x=0 で、確率は 50% となっています。これは標準正規分布が 0 を中心とした左右対称の分布だからです。



13 1)一様分布とは、全ての結果が同じ確率で生じる分布をいいます。ここで確率変数 X は連続確率変数であり、 a から b の間の値を一定の確率でとりうるとしています。下図は、密度関数を示したものです。



2)密度関数は高さが $1/(b-a)$ 、幅が $b-a$ となります。面積=高さ×幅ですから、面積が 1 となることを確認できます。

3) a から b の区間の midpoint $(a+b)/2$ が期待値となります。直観的には、期待値は分布の中心ですから、期待値が midpoint であることは理解しやすいでしょう。

厳密には、積分を用いて期待値を求めます。確率変数が離散的な値をとるのであれば、期待値は取りうる値を確率で加重平均することで求めることができました。連続的確率変数の場合も、期待値は、取りうる値を確率で加重平均することで求めることができます。ただし、その確率は密度関数 $f(x)$ であり、和 Σ は積分 $\int$ を用いることになります。数式で書くと、確率変数 X が $[a,b]$ 区間の値をとるとき、期待値は

$$E[X] = \int_a^b xf(x)dx$$

となります。

一様確率変数の場合、確率変数 X は、 $[a,b]$ 区間の範囲では $f(x)=1/(b-a)$ としましたから、その期待値は

$$E[X] = \int_a^b x \frac{1}{(b-a)} dx = \frac{1}{(b-a)} \int_a^b x dx$$

となります($1/(b-a)$ は定数ですから、 $\int$ 記号の外に出すことができます)。ここで上式に、

$$\int_a^b x dx = \left[\frac{x^2}{2} \right]_a^b = \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{(a+b)(a-b)}{2}$$

を代入すれば、 $E[X]=(a+b)/2$ を示せます。

第 7 章の答え

1. サンプルサイズは十分大きいため、中心極限定理を用いることができますし、標本分散は母分散を高い精度で推定できています。よって、平均 $\bar{X}$ は期待値 μ で分散 $3^2/50$ の正規分布に従います。

1) 平均と母平均 μ の差が 1 時間を越える確率は ($\sqrt{3^2/50}=0.424$ を用いて)、

$$P\{|\bar{X} - \mu| > 1\} = P\left\{\left|\frac{\bar{X} - \mu}{0.424}\right| > \frac{1}{0.424}\right\} = P\{|Z| > 2.36\}$$

となります。また、

$$\begin{aligned} P\{|Z| > 2.36\} &= P\{Z > 2.36\} + P\{Z < -2.36\} \\ &= 2 \times P\{Z > 2.36\} \\ &= 2 \times (1 - P\{Z < 2.36\}) \\ &= 2 \times (1 - 0.9909) = 0.0182 \end{aligned}$$

- 2) 標本平均を標準化すると標準正規分に従いますから、

$$P\left\{-1.96 < \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} < 1.96\right\} = 0.95$$

が成立します。したがって、 μ の 95%信頼区間は、

$$\bar{X} - 1.96 \sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + 1.96 \sqrt{\frac{\sigma^2}{n}}$$

となります。ここで、平均は 10 時間ですから、母平均 μ の 95%信頼区間は

$$10 - 1.96 \times 0.424 < \mu < 10 + 1.96 \times 0.424$$

です。これを計算すると $9.17 < \mu < 10.83$ となります。

2. 平均 $\bar{X}$ は期待値 μ で分散 $10^2/40$ の正規分布に従います。よって、95%信頼区間は、

$$450 - 1.96 \sqrt{\frac{10^2}{40}} < \mu < 450 + 1.96 \sqrt{\frac{10^2}{40}}$$

となります (これを計算すると $446.9 < \mu < 453.1$)。

3. サンプルサイズが十分に大きいとき、 $\hat{p} \sim N(p, p(1-p)/n)$ ですから

$$P\left\{-1.96 \sqrt{\frac{p(1-p)}{n}} < \hat{p} - p < 1.96 \sqrt{\frac{p(1-p)}{n}}\right\} = 0.95$$

となります。推定誤差が $\pm 1\%$ 以内に収まる確率は 95%、つまり

$$P\{-0.01 < \hat{p} - p < 0.01\} = 0.95$$

となります。よって、

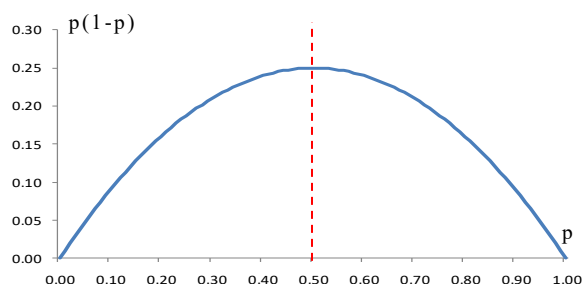
$$0.01 = 1.96 \sqrt{\frac{p(1-p)}{n}}$$

を解くことで、推定誤差が±1%以内に収まるために必要なサンプルサイズ n として、

$$n = 196^2 p(1-p)$$

を求めることができます。ここで n は p の値に依存しているため、 p の値が分かれば n を求めることが出来ます。しかし、調査前に p の値は未知ですから、 n を決めることはできません。

ここで $p(1-p)$ は、 $p=0.5$ で最大値になることを確認しましょう。下図は、横軸が p 、縦軸が $p(1-p)$ としています。 $p=0$ （もしくは1）なら $p(1-p)=0$ 、 $p=0.5$ なら $p(1-p)=0.25$ で最大となります¹。以上から、 $p=0.5$ のとき、 n は $196^2 \cdot 0.5^2 = 9604$ で最大となります。



以上から、たとえ最悪の状況（ $p=0.5$ ）を想定しても、9604 人に聞き取り調査をすれば、推定誤差は 1%以内に収まることが分かります。

4. 通常の調査では n は小さいため推定誤差は大きくなります。よって、0.1%の低下は単なる標本変動と考えられ、支持率が低下しているとはいえません。
5. 1) $p=300/N$ 、 $\hat{p}=30/200=0.15$ ですから、 $N=300/\hat{p}=300/0.15=2000$ 匹として推定されます。2) N の 95%信頼区間を求めるため、まずは p の 95%信頼区間を求めます。これは、

¹ $p(1-p)=p-p^2=-(p-1/2)^2+1/4$ と展開できます。ここで 1 項目だけが p に依存し、常に 0 以下となります。この項が最大となるのは 0 のとき、つまり $p=1/2$ のときです。

$$0.15 - 1.96\sqrt{\frac{0.15(1-0.15)}{200}} < p < 0.15 + 1.96\sqrt{\frac{0.15(1-0.15)}{200}}$$

となります（計算すると $0.10 < p < 0.20$ です）。よって、N の 95%信頼区間は下限が $300/0.20=1500$ 、上限が $300/0.10=3000$ となります。

6. $\hat{\theta}$ の分散は $V(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2] - E[\hat{\theta}]^2 = E[\hat{\theta}^2] - \theta^2$ ですから、 $V(\hat{\theta})=0$ でない限り、 $\hat{\theta}^2$ は θ^2 の不偏推定量ではありません²。

7. 平均 2 乗誤差は

$$\begin{aligned} E[(\hat{\theta} - \theta)^2] &= E\{(\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta)\}^2 \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta)^2 + 2(E[\hat{\theta}] - \theta)E(\hat{\theta} - E[\hat{\theta}]) \end{aligned}$$

と分解できます。ここで右辺第 3 項は

$$E(\hat{\theta} - E[\hat{\theta}]) = E[\hat{\theta}] - E[\hat{\theta}]$$

ですから 0 となります。また、右辺第 1 項は $\hat{\theta}$ の分散に他なりません。

平均 2 乗誤差とは？

教科書では紹介できませんでしたが、平均 2 乗誤差は、推定量が真の値と平均的にどの程度ずれるかを測っており、推定量の望ましさを測る重要な指標となっています。平均 2 乗誤差が小さいほど良い推定量とされ、これが大きいほど悪い推定量とされます。問題 7 から、平均 2 乗誤差は、推定量の分散とバイアスの 2 乗に分解されることが分かりました。

$$E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + (E[\hat{\theta}] - \theta)^2$$

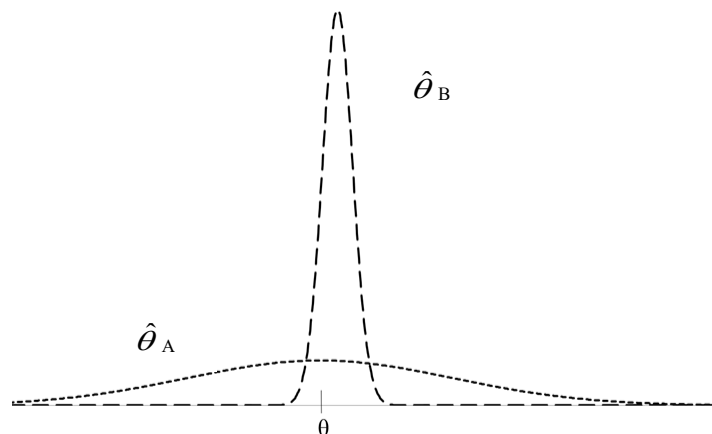
つまり、良い推定量とは、バイアスだけでなく、ばらつきも小さい推定量となるのです。

次の 2 つの推定量を考えてみてください(下図参照)。推定量 $\hat{\theta}_A$ は、バイアスは 0 ですが（不偏推定量）、分散が大きい推定量となっています。これに対し、推定量 $\hat{\theta}_B$ は、バイアスは少しありますが（不偏推定量ではない）、分散がとても小さいため、かなり正確に真の値を推定できます³。平均 2 乗誤差を計算すると、推定量 $\hat{\theta}_B$ の方が小さくなり、バイアスが存在したとし

² たとえば、 $\theta=0$ なら θ^2 も 0 ですが、 $\hat{\theta}$ が 0 を中心に分布しても、 $\hat{\theta}^2$ はプラスの値しかとりませんからその期待値は 0 ではありません。

³ ここでは両推定量とも一貫性は満たされるとしましょう。

でも、ばらつきが小さいため、より望ましい推定量といえます。この教科書では、不偏推定量だけを扱っていますが、よりアドバンスな統計学を勉強すると、バイアスがあっても分散が小さいなら、有用な推定量があることが分かります。



8. この推定量の期待値は⁴、

$$\begin{aligned} E\left[\frac{2}{n(n+1)}\sum_{i=1}^n iX_i\right] &= \frac{2}{n(n+1)}\sum_{i=1}^n iE[X_i] \\ &= \frac{2}{n(n+1)}\mu\sum_{i=1}^n i \\ &= \mu\frac{2}{n(n+1)}\frac{n(n+1)}{2} = \mu \end{aligned}$$

となりますから、不偏性が満たされています。

9. 1)期待値は以下となります。

$$E\left[\sum_{i=1}^n \varpi_i X_i\right] = \sum_{i=1}^n \varpi_i E[X_i] = \mu \sum_{i=1}^n \varpi_i$$

⁴式展開では、次の公式を用いました。

$$\sum_{i=1}^n i = 1 + 2 + 3 + \dots + (n-2) + (n-1) + n = \frac{n(n+1)}{2}$$

[証明]この公式を証明しましょう。ここで n が偶数とします。上式で、1 と n を足すと $n+1$ 、2 と $n-1$ を足すと $n+1$ 、3 と $(n-2)$ を足すと $n+1$ です。よって、 $\sum i$ は $n+1$ が $n/2$ 個ありますから、これらの和は $(n+1) \times n/2 = n(n+1)/2$ です。 n が奇数でも同様に証明できます。[終]

2) よって、 $\sum \omega_i = 1$ であれば不偏性が満たされます。つまり、加重平均は不推定量となります。

3) $\sum \omega_i = 1$ のとき、分散は

$$\begin{aligned} E\left[\left(\sum_{i=1}^n \omega_i X_i - \mu\right)^2\right] &= E\left[\left(\sum_{i=1}^n \omega_i X_i - \sum_{i=1}^n \omega_i \mu\right)^2\right] \\ &= E\left[\left(\sum_{i=1}^n \omega_i (X_i - \mu)\right)^2\right] \\ &= \sum_{i=1}^n \omega_i^2 E[(X_i - \mu)^2] = \sigma^2 \sum_{i=1}^n \omega_i^2 \end{aligned}$$

となります。式展開において、 $\sum_{i=1}^n \omega_i = 1$ から $\mu = \sum_{i=1}^n \omega_i \mu$ となること、 X_i は相互に独立であること⁵、を用いました。

この式から、分散が最小となるのは $\sum_{i=1}^n \omega_i^2$ が最小となるときである、と分かります。ここでは、分散を最小にする ω_i は、 $1/n$ であることを証明します。

$$\begin{aligned} \sum_{i=1}^n \omega_i^2 &= \sum_{i=1}^n \left(\left(\omega_i - \frac{1}{n} \right) + \frac{1}{n} \right)^2 = \sum_{i=1}^n \left(\omega_i - \frac{1}{n} \right)^2 + \frac{n}{n^2} + 2 \frac{1}{n} \sum_{i=1}^n \left(\omega_i - \frac{1}{n} \right) \\ &= \sum_{i=1}^n \left(\omega_i - \frac{1}{n} \right)^2 + \frac{1}{n} + 2 \frac{1}{n} \left(\sum_{i=1}^n \omega_i - \frac{n}{n} \right) \end{aligned}$$

右辺 2 項目は ω_i に依存しませんし、第 3 項は 0 となります(式展開において、 $\sum_{i=1}^n \omega_i = 1$ に注意)。よって、分散を最小にする ω_i は第 1 項目を 0 とする $1/n$ となります。したがって、平均が最も有効な推定量となります。

⁵ ここでは $n=2$ として、式展開をみていきましょう。

$$\begin{aligned} E\left[\left(\sum_{i=1}^2 \omega_i (X_i - \mu)\right)^2\right] &= E[(\omega_1(X_1 - \mu) + \omega_2(X_2 - \mu))^2] \\ &= E[\omega_1^2(X_1 - \mu)^2 + \omega_2^2(X_2 - \mu)^2 + 2\omega_1\omega_2(X_1 - \mu)(X_2 - \mu)] \\ &= \omega_1^2 E[(X_1 - \mu)^2] + \omega_2^2 E[(X_2 - \mu)^2] + 2\omega_1\omega_2 E[(X_1 - \mu)(X_2 - \mu)] \\ &= \omega_1^2 \sigma^2 + \omega_2^2 \sigma^2 + 2\omega_1\omega_2 \times 0 \\ &= \omega_1^2 \sigma^2 + \omega_2^2 \sigma^2 \end{aligned}$$

ここで X_1 と X_2 は独立ですから、共分散は 0 ($E[(X_1 - \mu)(X_2 - \mu)] = 0$) としました。

第 8 章の答え

1. 帰無仮説は棄却することに意義があります。教科書 8 章(p201-202)で詳しく説明しましたが、帰無仮説を採択しても、帰無仮説が正しいのか、対立仮説が正しいのか、どちらともいえません。
2. 有意水準は、状況に応じて設定を変えることが必要となります。有意水準を低く設定する場合としては、医薬品の開発やドーピング検査などが考えられます。医薬品開発では、帰無仮説（「医薬品の効果がない」）を誤って棄却して効果のない薬（または有害な薬）を市場に出すデメリットは大きいので有意水準は低く設定されます。また、ドーピング検査では、帰無仮説（「ドーピングをしていない」）を誤って棄却して選手のキャリアを傷つけるコストは大きいのでやはり有意水準は低く設定されます。逆に、有意水準を低く設定しないものとして、人間ドックの簡易検査があります。帰無仮説（「病気にかかっていない（陰性）」）を誤って棄却しても、精密検査を行えば良いだけですから大きな問題はありません。逆に、帰無仮説を誤って採択し、病気を放置してしまうコストは大きいといえます。
3. 仮説は $H_0: p=1/6$ （「サイコロに歪みがない」）、 $H_1: p \neq 1/6$ （「サイコロに歪みがある」）です。 H_0 が正しいもとで、 $\hat{p} \sim N(1/6, (1/6)(5/6)/300)$ となります（ $\sqrt{(1/6)(5/6)/300} = 0.0215$ ）。 $P\{1/6 - 1.96 \times 0.0215 < \hat{p} < 1/6 + 1.96 \times 0.0215\} = 0.95$ ですから、 $\hat{p}$ が $(0.1245, 0.2088)$ 区間に収まらなければ、有意水準 5% で帰無仮説は棄却されます。ここで、 $\hat{p} = 0.2$ ですから、 H_0 （「サイコロにゆがみはない」）は採択されます。
4. 仮説は $H_0: p=0.75 (=3/4)$ 、 $H_1: p \neq 0.75$ です。 H_0 が正しいもとで、 $\hat{p} \sim N(0.75, 0.25 \times 0.75 / 270)$ です。標準誤差は $\sqrt{0.25 \times 0.75 / 270} = 0.0264$ となります。ここで

$$P\{0.75 - 1.96 \times 0.0264 < \hat{p} < 0.75 + 1.96 \times 0.0264\} = 0.95$$
 から、 $\hat{p}$ が $(0.698, 0.802)$ 区間に収まらなければ、有意水準 5% で H_0 は棄却されます。ここで、 $\hat{p} = 200 / (200 + 70) = 0.741$ ですから、 $\hat{p}$ は区間内におさまっており、 H_0 は採択されます。よって、この結果から、メンデルの法則（「3 : 1 の割合で黄色と緑色のエンドウが生じる」）が誤っている、とは言

えません¹。

5. 仮説は $H_0: \mu=110$ 、 $H_1: \mu \neq 110$ です。 H_0 が正しいもとで、 $\bar{X} \sim N(110, 15^2/100)$ です ($\sqrt{15^2/100}=1.5$)。 $P\{110-1.96 \times 1.5 < \bar{X} < 110+1.96 \times 1.5\}=0.95$ から、 $\bar{X}$ が (107.06, 112.94) 区間に収まらなければ、有意水準 5% で H_0 は棄却されます。 $\bar{X}=120$ ですから、 H_0 は棄却されます。よって、今年度の新入生が受けたテスト結果は例年と同じとはいえません。

6. ここで仮説は $H_0: \mu_1 - \mu_2 = 0$ 、 $H_1: \mu_1 - \mu_2 \neq 0$ となります。 H_0 が正しいもとで、

$$\bar{X}_1 - \bar{X}_2 \sim N\left(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

となりますから、

$$P\{-1.96 \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \bar{X}_1 - \bar{X}_2 < 1.96 \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\} = 0.95$$

が成立します。サンプルサイズが十分に大きいため、分散 (σ_1^2 、 σ_2^2) は標本分散 (s_1^2 、 s_2^2) で置き換えることができます。ここで

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{200^2}{50} + \frac{150^2}{50}} = 35.35$$

ですから、 H_0 が正しいもとで、95%の確率で

$$-1.96 \times 35.35 < \bar{X}_1 - \bar{X}_2 < 1.96 \times 35.35$$

が成立します。もし $\bar{X}_1 - \bar{X}_2$ が区間内にあれば H_0 は採択され、区間外にあれば H_0 は棄却されます。平均差は $\bar{X}_1 - \bar{X}_2 = -300$ であり、これは $-1.96 \times 35.35 = -69.29$ より小さいため、 H_0 は棄却されます。よって、2種の時計の平均寿命は同じであるとはいえません。

¹ 読者の中には、 $H_0: p=0.25 (=1/4)$ 、 $H_1: p \neq 0.25$ と設定した方もおられるでしょう。この場合についても同じ結果が得られます。 H_0 が正しいもとで、 $\hat{p} \sim N(0.25, 0.25 \times 0.75/270)$ となります。ここで標準誤差は $\sqrt{0.25 \times 0.75/270} = 0.0263$ ですから、 $P\{0.25 - 1.96 \times 0.0263 < \hat{p} < 0.25 + 1.96 \times 0.0263\} = 0.95$ より、 $\hat{p}$ が (0.198, 0.302) 区間に収まらなければ有意水準 5% で H_0 は棄却されます。 $\hat{p} = 70/(200+70) = 0.259$ であり、区間内におさまっていますから H_0 は採択されます。よって、メンデルの法則が誤っているとは言えません。

第 9 章の答え

1. ここで

$$\frac{(5-1)s^2}{\sigma^2} \sim \chi^2(4)$$

から、

$$\begin{aligned} 0.95 &= P\left\{\chi_{4,0.975}^2 < \frac{(n-1)s^2}{\sigma^2} < \chi_{4,0.025}^2\right\} \\ &= P\left\{\frac{(n-1)s^2}{\chi_{4,0.025}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{4,0.975}^2}\right\} \end{aligned}$$

となります。 $\chi_{4,0.025}^2 = 11.14$ 、 $\chi_{4,0.975}^2 = 0.484$ から、 σ の 95%信頼区間は

$$\sqrt{\frac{(5-1)10^2}{11.14}} < \sigma < \sqrt{\frac{(5-1)10^2}{0.484}}$$

となります。これを計算すると、 $5.99 < \sigma < 28.75$ です。

2. $(10-1)s^2/\sigma^2 \sim \chi^2(9)$ ですから、95%の信頼区間は $\sqrt{(10-1)5^2/19.02} < \sigma < \sqrt{(10-1)5^2/2.70}$

でとなります。これを計算すると、 $3.44 < \sigma < 9.13$ です。

3. ここで

$$\frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{10}}} \sim t(9)$$

となりますから、

$$P\left\{-t_{9,0.05} < \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{10}}} < t_{9,0.05}\right\} = 0.95$$

が成立します。よって、 μ の 95%信頼区間は

$$\bar{X} - t_{9,0.05} \sqrt{\frac{s^2}{10}} < \mu < \bar{X} + t_{9,0.05} \sqrt{\frac{s^2}{10}}$$

となります。ここで標本平均は 52、標本分散は 75.78 ですから、これらを代入することで

$$52 - 2.262 \sqrt{\frac{75.78}{10}} < \mu < 52 + 2.262 \sqrt{\frac{75.78}{10}}$$

となります。これを計算すると、信頼区間は $45.77 < \mu < 58.23$ となります。

σ^2 に関する 95%信頼区間は $(10-1)75.78/19.02 < \sigma^2 < (10-1)75.78/2.70$ となります。これを計算すると、 $35.86 < \sigma^2 < 252.59$ ですから、 σ に関する信頼区間は $5.99 < \sigma < 15.89$ となります。

4. 95%信頼区間は $7 - 2.262\sqrt{3^2/10} < \mu < 7 + 2.262\sqrt{3^2/10}$ です ($4.85 < \mu < 9.15$)。
5. μ の 95%信頼区間は、 $0.01 - 2.145\sqrt{0.02^2/15} < \mu < 0.01 + 2.145\sqrt{0.02^2/15}$ です。これを計算すると $-0.001 < \mu < 0.021$ となります。
6. μ の 95%の信頼区間は $5 - 2.571\sqrt{2^2/6} < \mu < 5 + 2.571\sqrt{2^2/6}$ です (これは $2.901 < \mu < 7.099$)。
7. 標本分散 s^2 は、

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$= \frac{\sigma^2}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

と書き換えることができる。ここで s^2 の期待値は、

$$E[s^2] = \frac{\sigma^2}{n-1} E \left[\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \right]$$

となる ($\sigma^2/(n-1)$ は定数であるため、期待値の外にだせる)。さらに

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

であり、 χ^2 確率変数の期待値は自由度に等しいことに注意すると、

$$\frac{\sigma^2}{n-1} E \left[\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \right] = \frac{\sigma^2}{n-1} (n-1) = \sigma^2$$

となる。以上から、 s^2 は不偏性を満たした推定量といえる。

次に、 s^2 の分散は、

$$V(s^2) = E[(s^2 - \sigma^2)^2]$$

$$= E \left[\left(\frac{\sigma^2}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 - \frac{n-1}{n-1} \sigma^2 \right)^2 \right]$$

となります¹。[]内の第一項、第二項ともに $\sigma^2/(n-1)$ を含んでいるため、これを2乗してから外にだすと、

¹ 式展開において、 $\sigma^2 = \frac{n-1}{n-1} \sigma^2$ としています。

$$E\left[\left(\frac{\sigma^2}{n-1}\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2-\frac{n-1}{n-1}\sigma^2\right)^2\right]=\frac{\sigma^4}{(n-1)^2}E\left[\left(\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2-(n-1)\right)^2\right]$$

となります。ここで χ^2 確率変数の分散は、 $2 \times$ 自由度に等しいことに注意すると、上式は

$$\frac{\sigma^4}{(n-1)^2}E\left[\left(\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2-(n-1)\right)^2\right]=\frac{\sigma^4}{(n-1)^2}2(n-1)=\frac{2\sigma^4}{n-1}$$

となります²。

以上から、

$$V(s^2)=\frac{2\sigma^4}{n-1}$$

となります。この式をみると、 n が大きくなるにつれて分散は 0 に近づいていくことが分かります。不偏性が満たされ、 n が大きくなるにつれて分散は 0 に近づいていきますから、一貫性も満たされるわけです。

² 既に確認した通り、

$$\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$$

となります。 χ^2 確率変数の期待値はその自由度ですから、 $\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2$ の分散は

$$E\left[\left(\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2-E\left[\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2\right]\right)^2\right]=E\left[\left(\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2-(n-1)\right)^2\right]$$

となります。分散は $2 \times$ (自由度) に等しいため、これは $2(n-1)$ となります。

第 10 章の答え

1 最小 2 乗推定量の式は $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$ ですから、これを書き換えると $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$ となります。よって、 $X = \bar{X}$ のとき、回帰直線上の点は $Y = \bar{Y}$ となります。

2. Y の理論値の偏差 2 乗和

$$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

を書き換えましょう。p252 から理論値は

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

となり、練習問題 1 から

$$\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$$

となりますから、これらを偏差 2 乗和の式に代入すると

$$\begin{aligned} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 &= \sum_{i=1}^n (\hat{\alpha} + \hat{\beta}X_i - (\hat{\alpha} + \hat{\beta}\bar{X}))^2 \\ &= \sum_{i=1}^n (\hat{\beta}X_i - \hat{\beta}\bar{X})^2 \\ &= \hat{\beta}^2 \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

が得られます。この偏差 2 乗和を決定係数の式に代入すると、決定係数の別表現が得られます。

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{\hat{\beta}^2 \sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

この式は、決定係数を計算するので便利です (X と Y それぞれの偏差 2 乗和、 $\hat{\beta}$ だけで計算できる)。

さらに上式を書き換えてみましょう。 $\hat{\beta}$ の式を上式に代入すると、

$$\begin{aligned} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)^2 \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} &= \frac{(\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}))^2}{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} \\ &= \left(\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \right)^2 = \left(\frac{s_{XY}}{\sqrt{s_X^2 s_Y^2}} \right)^2 = r_{XY}^2 \end{aligned}$$

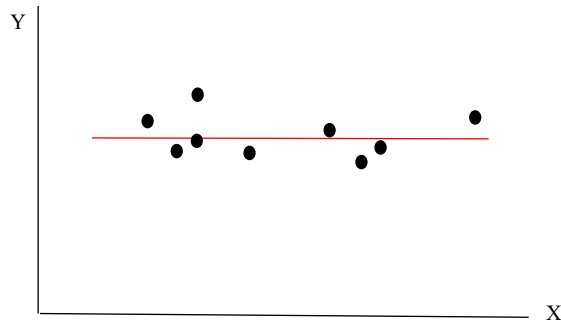
となります。つまり、決定係数とは、 X_i と Y_i との相関係数 r_{xy} を 2 乗したものに

他ならないのです。

3 (1)決定係数の別表現（練習問題 2 参照）から、 $\hat{\beta} = 0$ なら $R^2 = 0$ といえます。

$$R^2 = \hat{\beta}^2 \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 0 \times \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 0$$

直観的には、 $\hat{\beta} = 0$ なら、 X が変動しても Y の変動を全く説明できないため、 R^2 は 0 となるといえます(下図参照)。



(2) 決定係数が 0 であっても、 $\hat{\beta} = 0$ とは必ずしも言えません。たとえば、 X の変動に比べて、 Y の変動が非常に大きいなら、次の別表現から R^2 は 0 となります。

$$\hat{\beta}^2 \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

4 最小 2 乗推定量の式に、表で計算された値を代入すると、

$$\hat{\beta} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{-13.2}{1.2} = -11$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 11.6 - (-11 \times 0.4) = 16$$

となります。 β の推定値は-11 ですから、雨が降ると（ X が 1 になる）売り上げは 11 万円減ることになります。 α の推定値は 16 ですから、天気は晴れ（ $X=0$ ）であれば売り上げは 16 万円となります¹。2) 決定係数は、練習問題 2 の別表現を用いると、簡単に求めることができます。別表現は

¹この結果は次のようにも推定できます。雨の日は 2 日あり平均は(6+4)/2=5 万円、晴れた日は 3 日あり平均は(18+15+15)/3=16 万円です。よって、 α は 16、売り上げの差は 5-16=-11 から β は-11 と推定されます。

$$R^2 = \hat{\beta}^2 \frac{\sum (X_i - \bar{X})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = (-11)^2 \frac{1.2}{153.2}$$

ですから、これを計算すると 0.947 となります。よって、Y の全変動のうち 94.7%はモデルで説明できているといえます。

5. 残差の性質を用いて証明ができます（残差の性質は p263 参照）。まず、残差の和は 0 から、残差の平均も 0 となります。よって、残差と説明変数の標本共分散の分子は、

$$\sum_{i=1}^n (X_i - \bar{X}) \hat{u}_i = \sum_{i=1}^n X_i \hat{u}_i - \bar{X} \sum_{i=1}^n \hat{u}_i$$

となります。残差と説明変数の積和は 0 ですから、第 1 項は 0 となります。また、残差の和は 0 ですから第 2 項も 0 となります。以上から、標本共分散は 0 となり、標本相関係数も 0 となります。

6. 実証研究をしていると、係数の解釈をどのようにするかわからなくなることがあります。とくに変数の対数をとったり、とらなかつたりすると解釈がとたんに見えにくくなるものです。係数の解釈は実証分析において重要ですので、練習問題を解くことで学習してもらえたらと思います。

- 1) ここでモデルは $Y = \alpha + \beta X$ ですから、X が 1 単位増えることで Y は

$$(\alpha + \beta(X + 1)) - (\alpha + \beta X) = \beta$$

単位だけ変化します。例えば、 $\beta=2$ なら、X が 1 単位増えると Y は 2 単位増えることになります。

- 2) ここでモデルは $\ln(Y) = \alpha + \beta \ln(X)$ となります。このモデルでは、Y、X とともに対数をとるため、対数・対数(log-log)モデルと呼ばれます。ここで X が X' に変化したとき、Y は Y' に変化するとしみましょう。このとき、モデル式から、

$$\ln(Y') - \ln(Y) = (\alpha + \beta \ln(X')) - (\alpha + \beta \ln(X)) = \beta(\ln(X') - \ln(X))$$

が成立します。上式を β について解くことで

$$\beta = \frac{\ln(Y') - \ln(Y)}{\ln(X') - \ln(X)} = \frac{\ln\left(1 + \frac{Y' - Y}{Y}\right)}{\ln\left(1 + \frac{X' - X}{X}\right)} \approx \frac{\frac{Y' - Y}{Y}}{\frac{X' - X}{X}}$$

となります。したがって、 β は「X が 1% 変化したとき Y が何% 変化するか」、つまり、X に関する Y の弾力性となります。例えば、 $\beta=0.1$ なら、X が 1% 増え

ると Y は 0.1%増えることになります。

3) ここでモデルは $Y = \alpha + \beta \ln(X)$ です。このモデルでは、Y は対数をとらず、X だけ対数をとりますので、線形・対数(linear-log)モデルと呼ばれます。ここで X が X' に変化したとき、Yは Y' に変化するとします。このとき、モデル式から、

$$Y' - Y = (\alpha + \beta \ln(X')) - (\alpha + \beta \ln(X)) = \beta(\ln(X') - \ln(X))$$

が成立します。この関係式を β について解くことで

$$\beta = \frac{Y' - Y}{\ln(X') - \ln(X)} = \frac{Y' - Y}{\ln\left(1 + \frac{X' - X}{X}\right)} \approx \frac{Y' - Y}{\frac{X' - X}{X}}$$

となります。分母が X の変化率、分子が Y の変化量となります。上式は、

$$Y' - Y = \beta \frac{X' - X}{X}$$

と書き換えられますので、X が 1%変化（つまり、0.01）すると、Y は 0.01β 単位変化することがわかります。例えば、 $\beta=10$ なら X が 1%変化すると Y は $0.01 \times 10 = 0.1$ 単位だけ変化します。

4) ここでモデルは $\ln(Y) = \alpha + \beta X$ です。Y は対数、X は対数をとりませんから、対数・線形(log-linear)モデルと呼ばれます。Xが X' に変化したとき、Yは Y' に変化するとします。このとき、モデルの関係式から、

$$\ln(Y') - \ln(Y) = (\alpha + \beta X') - (\alpha + \beta X) = \beta(X' - X)$$

が成立します。この関係式を β について解くことで

$$\beta = \frac{\ln(Y') - \ln(Y)}{X' - X} = \frac{\ln\left(1 + \frac{Y' - Y}{Y}\right)}{X' - X} \approx \frac{\frac{Y' - Y}{Y}}{X' - X}$$

となります。上式は、

$$\frac{Y' - Y}{Y} = \beta(X' - X)$$

となります。X が X' へと 1 単位変化したら（つまり、 $X' - X = 1$ ）

$$\frac{Y' - Y}{Y} = \beta$$

となりますから、X の 1 単位の増加は Y を $100\beta\%$ だけ変化させます。例えば、 $\beta=0.01$ であれば、X の 1 単位の増加は Y を 1%増加させることがわかります。

第 11 章の答え

1. (1) β は真の値であり、 $\hat{\beta}$ はデータから推定される β の推定量です。

(2) モデルは $Y_i = \alpha + \beta X_i + u_i$ となるため、誤差項 u_i は

$$u_i = Y_i - \alpha - \beta X_i$$

となります。つまり、誤差項 u_i は「実現値 Y_i と真の回帰式からの理論値 $\alpha + \beta X_i$ との差」です。これに対して、残差 $\hat{u}_i$ は

$$\hat{u}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$$

となります。つまり、残差 $\hat{u}_i$ は「実現値 Y_i と推定された回帰式からの理論値 $\hat{\alpha} + \hat{\beta} X_i$ との差」です。誤差項 u_i の推定量が残差 $\hat{u}_i$ ともいえます。誤差項 u_i は観察できませんが、残差 $\hat{u}_i$ はデータから計算できるという違いもあります。

2. (1) 標準誤差は、以下となります。

$$\sqrt{s_{\hat{\beta}}^2} = \sqrt{\frac{\sum \hat{u}_i^2 / (n-2)}{\sum (X_i - \bar{X})^2}} = \sqrt{\frac{120/3}{250}} = 0.4$$

2) ここで $n=5$ ですから、 β の t 統計量は自由度 3 の t 分布に従います。よって、 β の信頼区間は、

$$\begin{aligned} 0.95 &= P\{-t_{3,0.05} < t_{\hat{\beta}} < t_{3,0.05}\} \\ &= P\left\{-t_{3,0.05} < \frac{\hat{\beta} - \beta}{\sqrt{s_{\hat{\beta}}^2}} < t_{3,0.05}\right\} \\ &= P\left\{\hat{\beta} - t_{3,0.05} \sqrt{s_{\hat{\beta}}^2} < \beta < \hat{\beta} + t_{3,0.05} \sqrt{s_{\hat{\beta}}^2}\right\} \end{aligned}$$

として求めることができます。ここで $t_{3,0.05} = 3.182$ から、信頼区間は $3.2 -$

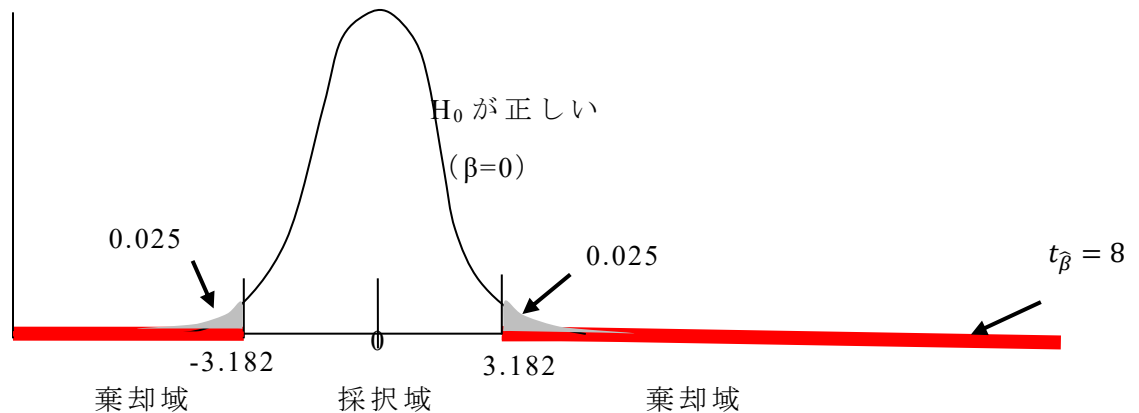
$$3.182 \times 0.4 < \beta < 3.2 + 3.182 \times 0.4$$

となります（計算すると $1.93 < \beta < 4.47$ です）。

(3) ここで t 値は

$$t_{\hat{\beta}} = \frac{\hat{\beta} - 0}{\sqrt{s_{\hat{\beta}}^2}} = \frac{3.2}{0.4} = 8$$

であり、これは $t_{3,0.05} = 3.182$ より大きいため H_0 を棄却します(下図参照)。



3. (1) $n=10$ 、 $t_{8,0.05} = 2.306$ から、95%の信頼区間は、

$$53.5 - 2.306 \times 5.3 < \beta < 53.5 + 2.306 \times 5.3$$

となります（計算すると $41.28 < \beta < 65.72$ ）。

(2) t 値は、

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta}{\sqrt{s_{\hat{\beta}}^2}} = \frac{53.5 - 55}{5.3} = -0.283$$

となり、 $|t_{\hat{\beta}}| = 0.283 < 2.306$ ですから H_0 を採択します。

(3) t 分布は左右対称ですから、片側検定を有意水準 5%で行うためには、 t 分布表の $\alpha=0.10$ の列から数値を見つけます。この場合、自由度 8 ですから、 $t_{\hat{\beta}}$ が -1.860 より小さいかで有意性を判断します¹。 t 値は $t_{\hat{\beta}} = -0.283$ であり、これは -1.860 より大きいため、 H_0 が採択されます。

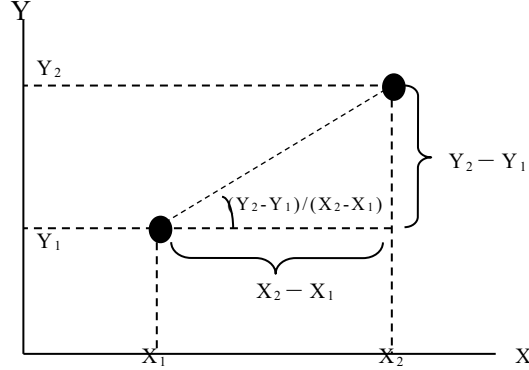
4. 1) 喫煙をしない場合 ($X=0$) 体重は 3kg、喫煙量が 1 日 20 本の場合 ($X=20$) 体重は $3 - 0.012 \times 20 = 2.76\text{kg}$ と予想されます。(2) サンプルサイズ n は 500 もあるので、 t 分布と正規分布はほとんど同じです。よって、 β の 95%信頼区間は、 $-0.012 - 1.96 \times 0.005 < \beta < -0.012 + 1.96 \times 0.005$ です ($-0.0218 < \beta < -0.0022$)。 (3)

$$t_{\hat{\beta}} = \frac{\hat{\beta}}{s_{\hat{\beta}}} = \frac{-0.012}{0.005} = -2.4$$

¹ 9 章で学習した通り、 t 分布表には、 $P\{t_{n,\alpha} < |U|\} = \alpha$ が成立する $t_{n,\alpha}$ の値が表されています。 t 分布は左右対称なので、 $P\{t_{n,\alpha} < U\} = P\{U < -t_{n,\alpha}\} = \alpha/2$ です。この問題では、 t 統計量が十分に小さいかで H_0 を採択するかが判断されます。よって、有意水準 α ならば、 t 統計量が $-t_{n,\alpha}$ より小さいかではなく（こうすると有意水準は $\alpha/2$ になる）、 $-t_{n,2\alpha}$ より小さいかで判断すればよいのです。

であり、 $|t_{\hat{\beta}}| = 2.4 > 1.96$ ですから有意水準 5%で H_0 を棄却します。

5. この推定量は何を意味しているか考えましょう。いま 2 個のデータ $((X_1, Y_1)$ と (X_2, Y_2)) があり、下図の●で表されています。



ここで、底辺 $X_2 - X_1$ 、高さ $Y_2 - Y_1$ となる三角形を考えましょう。この三角形の角度は $(Y_2 - Y_1) / (X_2 - X_1)$ で表されます。つまり、この推定量は、2 点から作られる三角形の角度に当たるわけです。この角度が、 X が 1 単位増えたとき、 Y が何単位変化するかを表すのは直観的でしょう。

この推定量の統計的性質をみてみましょう。ここでモデルは標準的仮定を満たしているとして、この推定量の期待値と分散を求めます。まず、この推定量の確率的表現を求めます。

$$\begin{aligned} \frac{Y_2 - Y_1}{X_2 - X_1} &= \frac{(\alpha + \beta X_2 + u_2) - (\alpha + \beta X_1 + u_1)}{X_2 - X_1} \\ &= \frac{\beta(X_2 - X_1) + (u_2 - u_1)}{X_2 - X_1} = \beta + \frac{u_2 - u_1}{X_2 - X_1} \end{aligned}$$

この期待値をとると、

$$\begin{aligned} E\left[\frac{Y_2 - Y_1}{X_2 - X_1}\right] &= E\left[\beta + \frac{u_2 - u_1}{X_2 - X_1}\right] \\ &= \beta + \frac{E[u_2] - E[u_1]}{X_2 - X_1} = \beta + \frac{0 - 0}{X_2 - X_1} = \beta \end{aligned}$$

となり、不偏性を満たしていることが確認できます。

次に、分散は、

$$E\left[\left(\frac{Y_2 - Y_1}{X_2 - X_1} - \beta\right)^2\right] = E\left[\left(\frac{u_2 - u_1}{X_2 - X_1}\right)^2\right] = \frac{E[(u_2 - u_1)^2]}{(X_2 - X_1)^2}$$

$$= \frac{E[u_1^2] + E[u_2^2] - 2E[u_1 u_2]}{(X_2 - X_1)^2} = \frac{2\sigma^2}{(X_2 - X_1)^2}$$

となります（標準的仮定から、 $E[u_i^2] = \sigma^2$ 、 $E[u_1 u_2] = 0$ としました）。

この推定量は 2 つの観測値しか使っていないため、最小 2 乗推定量より分散は大きくなっています²。分母は $(X_2 - X_1)^2$ ですから、 X_2 と X_1 が互いに離れているほど、その分散は小さくなり、より望ましい推定量となります（説明変数のばらつきが大きいほど正確に推定できる点については、p273 の図 11-4 を参照してください）。

6. この問題を通じて、変数のスケール変更によって推定結果に何が生じるかを理解しましょう。もともと X_i と Y_i には $Y_i = \alpha + \beta X_i + u_i$ という関係があります（ u の分散は σ^2 ）。この式の両辺に c_y を掛けると

$$c_y Y_i = c_y \alpha + \left(\frac{c_y}{c_x} \beta \right) c_x X_i + c_y u_i$$

と書き換えられます。ここで $Y_i^* = c_y Y_i$ 、 $X_i^* = c_x X_i$ 、 $\alpha^* = c_y \alpha$ 、 $\beta^* = (c_y / c_x) \beta$ 、 $u_i^* = c_y u_i$ と定義すると、上式は

$$Y_i^* = \alpha^* + \beta^* X_i^* + u_i^*$$

と表記できます（ u_i^* の分散を σ^{2*} と表記すれば $\sigma^{2*} = E[(c_y u_i)^2] = c_y^2 \sigma^2$ ）。

この式から、スケール変更した変数 Y^* 、 X^* 同士にどのような関係があるかを

² 直観的に自明ですが、最小 2 乗推定量の分散の方が小さいこと、つまり、

$$\frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} < \frac{2\sigma^2}{(X_2 - X_1)^2}$$

を証明します。具体的には、上式を書き換えた次式を証明します。

$$(X_2 - X_1)^2 < 2 \sum_{i=1}^n (X_i - \bar{X})^2$$

【証明】 第 1 項と第 2 項はともに 2 乗和であるため、次式が成立します。

$$0 < \{(X_2 - \bar{X}) + (X_1 - \bar{X})\}^2 + 2 \sum_{i=3}^n (X_i - \bar{X})^2$$

ここで、右辺の第 2 項の下添字は 3 としています。ここで、 $\{(X_2 - \bar{X}) + (X_1 - \bar{X})\}^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + 2(X_1 - \bar{X})(X_2 - \bar{X})$ に注意すると、上式は、

$$-2(X_1 - \bar{X})(X_2 - \bar{X}) < (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + 2 \sum_{i=3}^n (X_i - \bar{X})^2$$

と書き換えられます。両辺に $(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2$ を足すと、

$$\underbrace{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 - 2(X_1 - \bar{X})(X_2 - \bar{X})}_{=(X_2 - X_1)^2} < \underbrace{2[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2] + 2 \sum_{i=3}^n (X_i - \bar{X})^2}_{2 \sum_{i=1}^n (X_i - \bar{X})^2}$$

左辺は、

$$(X_2 - X_1)^2 = \{(X_2 - \bar{X}) - (X_1 - \bar{X})\}^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 - 2(X_1 - \bar{X})(X_2 - \bar{X})$$

であり、右辺は $2 \sum_{i=1}^n (X_i - \bar{X})^2$ となります。 [終]

理解できるでしょう。たとえば、説明変数 X だけを 100 倍したとしましょう ($c_y=1$ 、 $c_x=100$)。このとき、新しい説明変数 X^* の係数 β^* は、 β を 1/100 倍したものとなります。これに対し、被説明変数だけを 100 倍したら ($c_y=100$ 、 $c_x=1$)、説明変数 X^* (この場合、 $X^*=X$) の係数 β^* は、 β を 100 倍したものとなります。以上から、スケール変更によって、推定されるパラメータの値が影響を受けることが分かりました。

以下では、スケール変更しても、t 統計量の値は全く影響を受けないことを示します。 β^* の最小 2 乗推定量は、

$$\begin{aligned}\hat{\beta}^* &= \frac{\sum_{i=1}^n (X_i^* - \bar{X}^*)(Y_i^* - \bar{Y}^*)}{\sum_{i=1}^n (X_i^* - \bar{X}^*)^2} \\ &= \frac{\sum_{i=1}^n (c_x X_i - c_x \bar{X})(c_y Y_i - c_y \bar{Y})}{\sum_{i=1}^n (c_x X_i - c_x \bar{X})^2} \\ &= \frac{c_x c_y}{c_x^2} \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{c_y}{c_x} \hat{\beta}\end{aligned}$$

と書き換えることができます。当然ですが、この推定量は不偏性を満たしています ($E[\hat{\beta}^*] = E[(c_y/c_x)\hat{\beta}] = (c_y/c_x)E[\hat{\beta}] = (c_y/c_x)\beta = \beta^*$)。次に、 $\hat{\beta}^*$ の分散は

$$s_{\hat{\beta}^*}^2 = \frac{\sigma^{*2}}{\sum_{i=1}^n (X_i^* - \bar{X}^*)^2}$$

となります。上式に、 $\sigma^{*2} = c_y^2 \sigma^2$ 、 $X_i^* = c_x X_i$ 、また

$$\bar{X}^* = \frac{\sum_{i=1}^n X_i^*}{n} = \frac{\sum_{i=1}^n c_x X_i}{n} = c_x \frac{\sum_{i=1}^n X_i}{n} = c_x \bar{X}$$

を代入すると、

$$s_{\hat{\beta}^*}^2 = \frac{c_y^2 \sigma^2}{\sum_{i=1}^n (X_i^* - \bar{X}^*)^2} = \frac{c_y^2 \sigma^2}{\sum_{i=1}^n (c_x X_i - c_x \bar{X})^2} = \frac{c_y^2}{c_x^2} \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

と書き換えることができます。以上から、t 統計量は

$$\frac{\hat{\beta}^* - \beta^*}{\sqrt{s_{\hat{\beta}^*}^2}} = \frac{\frac{c_y}{c_x} \hat{\beta} - \frac{c_y}{c_x} \beta}{\sqrt{\frac{c_y^2}{c_x^2} \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}} = \frac{\frac{c_y}{c_x} (\hat{\beta} - \beta)}{\frac{c_y}{c_x} \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}}$$

$$= \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}} = \frac{\hat{\beta} - \beta}{\sqrt{s_{\hat{\beta}}^2}}$$

となります（式展開では、 $\hat{\beta}^* = (c_y/c_x)\hat{\beta}$ 、 $\beta^* = (c_y/c_x)\beta$ に注意）。この結果から、スケール変更により、t 統計量は影響を受けないことが確認できました。

7. 以下では、最小 2 乗推定量 $\hat{\alpha}$ と $\hat{\beta}$ の確率的表現を用いるので再掲載します。

$$\begin{aligned}\hat{\alpha} &= \alpha - (\hat{\beta} - \beta)\bar{X} + \bar{u} \\ \hat{\beta} &= \beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2}\end{aligned}$$

ここで $\hat{\alpha}$ の分散は

$$\begin{aligned}E[(\hat{\alpha} - \alpha)^2] &= E[(-(\hat{\beta} - \beta)\bar{X} + \bar{u})^2] \\ &= \bar{X}^2 E[(\hat{\beta} - \beta)^2] - 2\bar{X} E[(\hat{\beta} - \beta)\bar{u}] + E[\bar{u}^2]\end{aligned}$$

と 3 つに分解できます。第 1 項は、 $\hat{\beta}$ の分散式である $\sigma_{\hat{\beta}}^2 = \sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2$ を用いて計算できます。第 2 項は、

$$\begin{aligned}E[(\hat{\beta} - \beta)\bar{u}] &= E\left[\left(\frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2}\right) \left(\frac{\sum_{i=1}^n u_i}{n}\right)\right] \\ &= \frac{1}{n \sum_{i=1}^n (X_i - \bar{X})^2} E\left[\left(\sum_{i=1}^n (X_i - \bar{X}) u_i\right) \left(\sum_{i=1}^n u_i\right)\right] = 0\end{aligned}$$

を用いると 0 となることを示せます³。第 3 項は

$$\begin{aligned}E[\bar{u}^2] &= E\left[\left(\frac{\sum_{i=1}^n u_i}{n}\right)^2\right] \\ &= \frac{1}{n^2} E[\sum_{i=1}^n u_i^2 + \sum_{i \neq j} \sum u_i u_j] = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

となります（ $i \neq j$ のとき $E[u_i u_j] = 0$ を用いました）。以上から、

$$E[(\hat{\alpha} - \alpha)^2] = \sigma^2 \left(\frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{1}{n} \right)$$

と書けます。この式から、n が大きくなるにつれて、（両項とも分母が無限大に

³ 期待値の部分が 0 となることを n=2 として証明しましょう。

$$\begin{aligned}E\left[\left(\sum_{i=1}^2 (X_i - \bar{X}) u_i\right) \left(\sum_{i=1}^2 u_i\right)\right] &= E[(X_1 - \bar{X}) u_1 + (X_2 - \bar{X}) u_2] (u_1 + u_2) \\ &= (X_1 - \bar{X}) E[u_1^2] + (X_2 - \bar{X}) E[u_2^2] + (X_1 - \bar{X}) E[u_1 u_2] + (X_2 - \bar{X}) E[u_1 u_2] \\ &= (X_1 - \bar{X}) \sigma^2 + (X_2 - \bar{X}) \sigma^2 = \sigma^2 \sum_{i=1}^2 (X_i - \bar{X}) = 0\end{aligned}$$

式展開において、 $E[u_1^2] = \sigma^2$ 、 $E[u_2^2] = \sigma^2$ 、 $E[u_1 u_2] = 0$ 、また偏差の和は 0 を用いました。

大きくなるため) 分散が 0 に近づくのは明らかでしょう。

分散は次のようにも書き換えられます。

$$\begin{aligned}
 E[(\hat{\alpha} - \alpha)^2] &= \frac{\sigma^2 \bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{\sigma^2}{n} \\
 &= \frac{\sigma^2 n \bar{X}^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} + \frac{\sigma^2 \sum_{i=1}^n (X_i - \bar{X})^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} \\
 &= \frac{\sigma^2 n \bar{X}^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} + \sigma^2 \frac{\sum_{i=1}^n X_i^2 - n \bar{X}^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}
 \end{aligned}$$

練習問題 8 次式が正しいことを証明せよ(補足参照)。

$$E[(\sum_{i=1}^n (X_i - \bar{X}) u_i)^2] = \sigma^2 \sum_{i=1}^n (X_i - \bar{X})^2$$

これは 2023 年以降に出版された本に含まれた問題です。教科書に練習問題 8 がない方も試しに解いてみてください。

[練習問題 8 の答え]

11 章補足では、証明が書かれていますが、より詳しい証明を記載します。まず、表記を簡潔にするため、 $v_i = (X_i - \bar{X}) u_i$ と定義します。次に、この定義を用いて、上式の左辺を次のように表記します。

$$E[(\sum_{i=1}^n (X_i - \bar{X}) u_i)^2] = E[(\sum_{i=1}^n v_i)^2]$$

ここで右辺を展開すると、次のようになります([]内を見やすくするため、右辺は n 行 n 列で配置しました)。

$$\begin{aligned}
 E \left[\left(\sum_{i=1}^n v_i \right)^2 \right] &= E[(v_1 + v_2 + v_3 + \dots + v_n)(v_1 + v_2 + v_3 + \dots + v_n)] \\
 &= E \begin{bmatrix} v_1^2 + v_1 v_2 + v_1 v_3 + v_1 v_4 + \dots + v_1 v_n + \\ v_2 v_1 + v_2^2 + v_2 v_3 + v_2 v_4 + \dots + v_2 v_n + \\ v_3 v_1 + v_3 v_2 + v_3^2 + v_3 v_4 + \dots + v_3 v_n + \\ \dots \\ v_n v_1 + v_n v_2 + v_n v_3 + v_n v_4 + \dots + v_n^2 \end{bmatrix}
 \end{aligned}$$

各項の期待値をとると、次のようになります。

$$\begin{bmatrix} E[v_1^2] + E[v_1v_2] + E[v_1v_3] + E[v_1v_4] + \dots + E[v_1v_n] + \\ E[v_1v_2] + E[v_2^2] + E[v_2v_3] + E[v_2v_4] + \dots + E[v_2v_n] + \\ E[v_1v_3] + E[v_2v_3] + E[v_3^2] + E[v_3v_4] + \dots + E[v_3v_n] + \\ \dots \\ E[v_1v_n] + E[v_2v_n] + E[v_3v_n] + E[v_4v_n] + \dots + E[v_n^2] \end{bmatrix}$$

ここで、標準的仮定 3($E[u_i^2] = \sigma^2$)から、

$$E[v_i^2] = E[(X_i - \bar{X})u_i]^2 = (X_i - \bar{X})^2 E[u_i^2] = (X_i - \bar{X})^2 \sigma^2$$

となること、また、標準的仮定 4($i \neq j$ なら $E[u_i u_j] = 0$)から、

$$E[v_i v_j] = E[(X_i - \bar{X})u_i (X_j - \bar{X})u_j] = (X_i - \bar{X})(X_j - \bar{X}) E[u_i u_j] = (X_i - \bar{X})(X_j - \bar{X}) \times 0 = 0$$

となることに注意してください。これらの関係式を用いると、次のようになります。

$$\begin{bmatrix} E[v_1^2] + E[v_1v_2] + E[v_1v_3] + E[v_1v_4] + \dots + E[v_1v_n] + \\ E[v_1v_2] + E[v_2^2] + E[v_2v_3] + E[v_2v_4] + \dots + E[v_2v_n] + \\ E[v_1v_3] + E[v_2v_3] + E[v_3^2] + E[v_3v_4] + \dots + E[v_3v_n] + \\ \dots \\ E[v_1v_n] + E[v_2v_n] + E[v_3v_n] + E[v_4v_n] + \dots + E[v_n^2] \end{bmatrix} = \begin{bmatrix} \sigma^2(X_1 - \bar{X})^2 + 0 + 0 + 0 + \dots + 0 + \\ 0 + \sigma^2(X_2 - \bar{X})^2 + 0 + 0 + \dots + 0 + \\ 0 + 0 + \sigma^2(X_3 - \bar{X})^2 + 0 + \dots + 0 + \\ \dots \\ 0 + 0 + 0 + 0 + \dots + \sigma^2(X_n - \bar{X})^2 \end{bmatrix}$$

ここで、左辺の対角線上の要素は $(X_i - \bar{X})^2 \sigma^2$ 、それ以外はすべて 0 としました。
これはまさに、

$$\sigma^2[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + \dots + (X_n - \bar{X})^2]$$

であり、 Σ 記号を用いて表記すると次式になります。

$$\sigma^2 \sum_{i=1}^n (X_i - \bar{X})^2$$

第 12 章の答え

練習問題 1

- (1)それぞれの t 統計量を計算すると、 $-0.1/0.05=-2$ 、 $0.15/0.04=3.75$ 、 $0.2/0.08 = 2.5$ です。サンプルサイズ 500 もあれば、 t 分布は標準正規分布とほぼ同じですから、有意性は t 値の絶対値が 1.96 より大きいかで判断します。ここで、 -2 、 3.75 、 2.5 は絶対値で 1.96 より大きいため、係数 0 の帰無仮説は棄却されます。
- (2)兄弟の数が増えると、自分以外の兄弟への教育費などがかさみ、家計に財政的余裕がなくなると考えられます。これが兄弟の数が増えると、教育年数が減ってしまう原因かもしれません。
- (3)父母の教育年数が高いと、彼らの所得も高い傾向にあるでしょうから、家計に財政的な余裕が生まれ、教育により多くの出費を厭わなくなる可能性があります。また、教育年数が高い親ほど、子供の教育により熱心な傾向があるのかもしれない。
- (4)太郎君の教育年数は $10-0.1 \times 2 + 0.15 \times 16 + 0.2 \times 12 = 14.6$ 年と予測されます。

練習問題 2

睡眠時間を除いて回帰分析をしてみましょう。

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

このとき、 $1 = (X_1 + X_2 + X_3 + X_4)/24$ ですから、

$$Y = \alpha \left(\frac{1}{24} X_1 + \frac{1}{24} X_2 + \frac{1}{24} X_3 + \frac{1}{24} X_4 \right) + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

となります。上式を展開すると、

$$Y = \left(\beta_1 + \frac{\alpha}{24} \right) X_1 + \left(\beta_2 + \frac{\alpha}{24} \right) X_2 + \left(\beta_3 + \frac{\alpha}{24} \right) X_3 + \frac{\alpha}{24} X_4 + u$$

となります¹。したがって、睡眠時間 X_4 の効果は $\alpha/24$ として測られます²。

¹ そもそもモデルを定数項なしの式で推定したら、結果の解釈は容易かもしれません。つまり、モデルを $Y = \beta_1^* X_1 + \beta_2^* X_2 + \beta_3^* X_3 + \beta_4^* X_4$ とします ($\beta_1^* = \beta_1 + \alpha/24$ 、 $\beta_2^* = \beta_2 + \alpha/24$ 、 $\beta_3^* = \beta_3 + \alpha/24$ 、 $\beta_4^* = \alpha/24$)。係数 β_i^* は、 X_i が 1 単位変化したとき Y が何単位変化するかを表します。

² この結果から、 β_i は、 X_i と X_4 との効果の差を測っていることが分かります。たとえば、 X_1 の Y への効果は $\beta_1 + \alpha/24$ 、 X_4 の Y への効果は $\alpha/24$ ですから、両者の差は β_1 となるわけです。

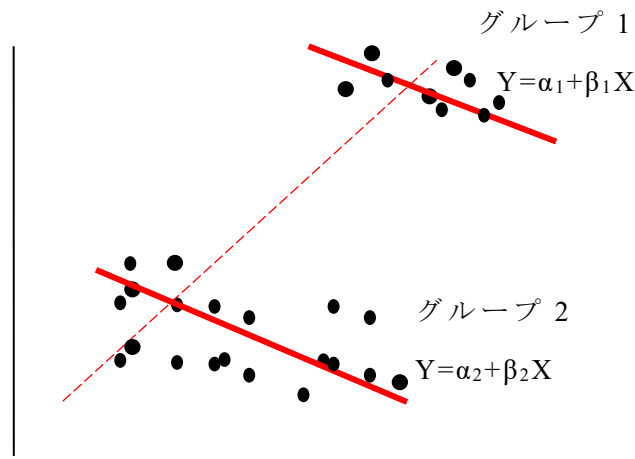
練習問題 3

双子は能力や家庭環境などにほとんど差がないと考えられますから、双子の変数の差を考えると、能力や家庭環境の影響などを取り除くことができます。したがって、Y を X で回帰すれば、バイアスなく X から Y への影響を測ることができます。これが双子のデータを用いる利点です。

しかしながら、双子のデータを用いれば何の問題もなく、X から Y への影響を正確に推定できるわけではありません。現実問題として、双子であれば教育年数にほとんど違いはないでしょう。つまり、双子の教育年数差 X はほぼ 0 の値をとり、(説明変数の変動が非常に小さいため) 推定結果は不安定となってしまいます。これが双子のデータを用いる欠点です。

練習問題 4

様々な原因が考えられますが、1 つの例を紹介します。下図をみると、各グループとも、係数 β はマイナスとなっています。しかし、グループを区別しないで一緒にすると、推定された β (図の点線の傾き) はプラスとなります。



練習問題 5

以下のような 11 個のダミー変数を定義します。

$$M1_t = \begin{cases} 1 & 1 \text{ 月} \\ 0 & \text{その他} \end{cases}, \quad M2_t = \begin{cases} 1 & 2 \text{ 月} \\ 0 & \text{その他} \end{cases}, \quad \dots, \quad M11_t = \begin{cases} 1 & 11 \text{ 月} \\ 0 & \text{その他} \end{cases}$$

12 個のダミー変数を定義して、説明変数として用いてしまうと、多重共線性が生じてしまうことに注意してください。

練習問題 6

1) 恒等式は、変数間の関係が分かっているため、そもそも式を推定する必要はありません。恒等式の推定は、何の意味もありません。恒等式の分析は、データ分析の初心者にはありがちなミスなので気をつけてください。

2) GDP の定義から、

$$\text{GDP} = \text{消費} + \text{投資} + \text{政府支出} + \text{輸出} - \text{輸入}$$

となります（これはまさに恒等式）。ここで被説明変数を GDP、説明変数を消費、投資、政府支出、輸出、輸入としましょう。このとき、消費、投資、政府支出、輸出の係数は 1、輸入の係数は -1 と推定されます。加えて、決定係数は 1 となります（恒等式ですから、当てはまりは 100% です）。しかし、GDP は定義から、上記の結果が得られることは自明であり、そもそも推定する必要などはないでしょう。

練習問題 7

モデルは、 $Y_{i,t} = \alpha + \beta X_{i,t} + \gamma_1 Z_i + \gamma_2 S_t + u_{i,t}$ ($i=1, \dots, N, t=1, \dots, T$) となります。 Z_i は時間を通じて一定な i 固有の要因（固定効果）、 S_t は時間を通じて変化し全体に同じ影響を与える要因（時間効果）です。たとえば、 S_t は景気変動を捉えるマクロ変数かもしれません。このモデルは、 $\alpha_i = \alpha + \gamma_1 Z_i$ 、 $\omega_t = \gamma_2 S_t$ と定義すると、

$$Y_{i,t} = \beta X_{i,t} + \alpha_i + \omega_t + u_{i,t}$$

と表現できます。 α_i は固定効果、 ω_t は時間効果を捉えています。ここで ω_t は、時間を通じて変化する要因をコントロールしています。

時間効果をコントロールするため、以下のダミー変数を定義しましょう。

$$T2_t = \begin{cases} 1 & t=2 \\ 0 & \text{その他} \end{cases}, \quad T3_t = \begin{cases} 1 & t=3 \\ 0 & \text{その他} \end{cases}, \quad TT_t = \begin{cases} 1 & t=T \\ 0 & \text{その他} \end{cases}, \dots,$$

上記のダミー変数は、時間 t だけに依存するため（ i に依存しない）、 i の添字を除いています。これらの変数だけでなく、固定効果をコントロールするため、以下のダミー変数を定義します(12 章参照)。

$$D2_i = \begin{cases} 1 & i=2 \\ 0 & \text{その他} \end{cases}, \quad D3_i = \begin{cases} 1 & i=3 \\ 0 & \text{その他} \end{cases}, \quad Dn_i = \begin{cases} 1 & i=n \\ 0 & \text{その他} \end{cases}, \dots,$$

このとき、先の式は

$$Y_{i,t} = \beta X_{i,t} + c + \theta_2 D2_t + \dots + \theta_n Dn_t + \eta_2 T2_t + \dots + \eta_T TT_t + u_{i,t}$$

と表現できます。ただし、パラメータは $c = \alpha_1 + \omega_1$ 、 $\theta_2 = \alpha_2 - \alpha_1$ 、 $\dots$ 、 $\theta_n = \alpha_n - \alpha_1$ 、 $\eta_2 = \omega_2 - \omega_1$ 、 $\dots$ 、 $\eta_T = \omega_T - \omega_1$ と定義します。ここで $i=1$ 、 $t=1$ なら全てのダミー変数が 0 となりますから、上式は

$$\begin{aligned} Y_{i,t} &= \beta X_{i,t} + c + u_{i,t} \\ &= \beta X_{i,t} + \alpha_1 + \omega_1 + u_{i,t} \end{aligned}$$

となります。また、 $i=1$ 、 $t=2$ なら $T2$ 以外のダミー変数が 0 となりますから

$$\begin{aligned} Y_{i,t} &= \beta X_{i,t} + c + \eta_2 + u \\ &= \beta X_{i,t} + (\alpha_1 + \omega_1) + (\omega_2 - \omega_1) + u_i \\ &= \beta X_{i,t} + \alpha_1 + \omega_2 + u_{i,t} \end{aligned}$$

となります。同じ i であっても、1 期と 2 期で ω の値が異なりますから、時間効果が考慮されていることが分かります。

練習問題 8

1) 最小 2 乗法は、残差 2 乗和である

$$\sum_{i=1}^n \tilde{u}_i^2 = \sum_{i=1}^n (Y_i - \tilde{\beta} X_i)^2$$

を最小化するように $\tilde{\beta}$ を選びます。上式を $\tilde{\beta}$ で微分して 0 と置くと、

$$-2 \sum_{i=1}^n (Y_i - \hat{\beta}_r X_i) X_i = 0$$

となります。この式を満たす $\tilde{\beta}$ は最小 2 乗推定量なので $\hat{\beta}_r$ と表記しました(下添字の r は、 $\alpha = 0$ という制約(restriction)を置いた推定であることを明示しています)。さらに両辺を -2 で割ると、以下の正規方程式が得られます。

$$\sum_{i=1}^n (Y_i - \hat{\beta}_r X_i) X_i = 0$$

さらに整理すると、

$$\sum_{i=1}^n X_i Y_i - \hat{\beta}_r \sum_{i=1}^n X_i^2 = 0$$

となり、この式を $\hat{\beta}_r$ について解くと、以下の最小 2 乗推定量が得られます。

$$\hat{\beta}_r = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$$

この式は、これまで OLS 推定量の式と異なることは明らかでしょう。

2) 真のモデルは $Y_i = \alpha + \beta X_i + u_i$ とします。 $\hat{\beta}_r$ の式において、 Y_i を $\alpha + \beta X_i + u_i$ で置き換えることで $\hat{\beta}_r$ の確率的表現を得ることができます。

$$\begin{aligned}\hat{\beta}_r &= \frac{\sum_{i=1}^n X_i (\alpha + \beta X_i + u_i)}{\sum_{i=1}^n X_i^2} = \frac{\alpha \sum_{i=1}^n X_i + \beta \sum_{i=1}^n X_i^2 + \sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2} \\ &= \beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} + \frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2}\end{aligned}$$

確率的表現の期待値をとると、

$$\begin{aligned}E[\hat{\beta}_r] &= \beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} + \frac{\sum_{i=1}^n X_i E[u_i]}{\sum_{i=1}^n X_i^2} \\ &= \frac{\alpha \sum_{i=1}^n X_i + \beta \sum_{i=1}^n X_i^2 + 0}{\sum_{i=1}^n X_i^2} \\ &= \beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2}\end{aligned}$$

となります。ここで $\alpha = 0$ (真のモデルに定数項がない) なら、2 項目は 0 になりますから、推定量は不偏性を満たします。しかし、 $\alpha \neq 0$ (真のモデルに定数項がある) なら、 $(\sum_{i=1}^n X_i = 0 \text{ ではない限り})$ 2 項目は 0 ではないため、推定量は不偏性を満たしません。当然ですが、誤った制約を課して推定してしまうとバイアスが発生してしまうことが理解できます。

上の関係式から、 $\sum_{i=1}^n X_i > 0$ であれば、 $\alpha > 0$ なら

$$E[\hat{\beta}_r] = \beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} > \beta$$

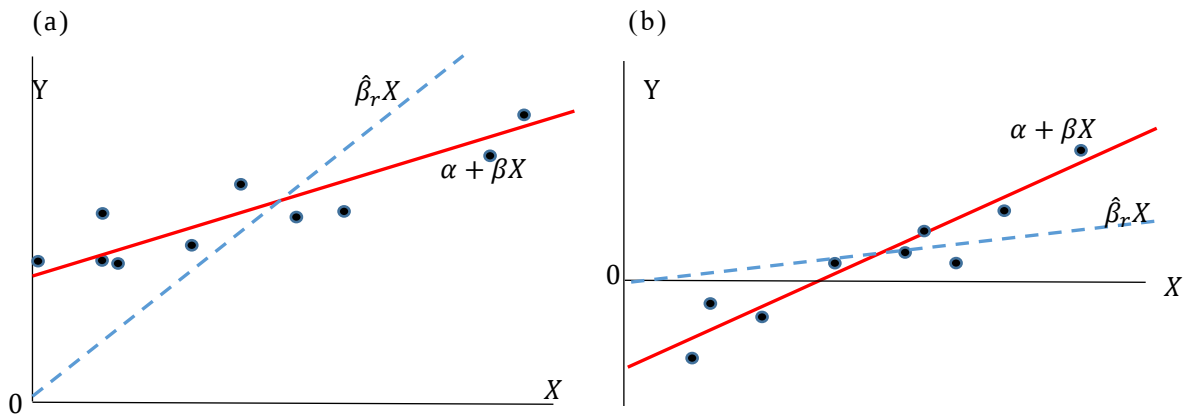
となり、逆に $\alpha < 0$ なら

$$E[\hat{\beta}_r] = \beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} < \beta$$

となります。これはなぜでしょうか。図を使って考えてみましょう。

下図では、 X_i は全て非負であるとします。実線が真の関係を表す式であり、点線が回帰式となります。図をみると、実線の周りでデータが観察されていることがわかります。 $\alpha=0$ と制約を置いて β を推定したとき、回帰直線は原点を必ず通ることになります。下図(a)では、 $\alpha > 0$ と仮定しており、図からも $\hat{\beta}_r$ が β より大きいことがわかります。これに対し、(b)では、 $\alpha < 0$ と仮定しており、図から $\hat{\beta}_r$ が β より小さいことがわかります。これらの図から、 α の符号によってバイアスの方向が変わるのがわかって頂けたかと思います。

図 バイアスと α の関係について



3)定数項がないとき、理論値と残差は

$$\hat{Y}_i = \hat{\beta}_r X_i, \quad \hat{u}_i = Y_i - \hat{\beta}_r X_i$$

となります。また、残差が満たすべき性質は、

$$\sum_{i=1}^n X_i (Y_i - \hat{\beta}_r X_i) = \sum_{i=1}^n X_i \hat{u}_i = 0$$

だけであり、この性質から、理論値と残差の積和は0となります。

$$\sum_{i=1}^n \hat{Y}_i \hat{u}_i = \sum_{i=1}^n (\hat{\beta}_r X_i) \hat{u}_i = \hat{\beta}_r \sum_{i=1}^n X_i \hat{u}_i = 0$$

しかし、残差の和が0という性質は満たされません。

以上から、Yの全変動は、

$$\begin{aligned} \sum_{i=1}^n (Y_i - \bar{Y})^2 &= \sum_{i=1}^n (\hat{Y}_i + \hat{u}_i - \bar{Y})^2 = \sum_{i=1}^n ((\hat{Y}_i - \bar{Y}) + \hat{u}_i)^2 \\ &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{u}_i^2 + 2 \sum_{i=1}^n (\hat{Y}_i - \bar{Y}) \hat{u}_i \\ &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{u}_i^2 - 2 \bar{Y} \sum_{i=1}^n \hat{u}_i \end{aligned}$$

となります（式展開では $\sum_{i=1}^n \hat{Y}_i \hat{u}_i = 0$ を用いました）。つまり、

$$\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} + \frac{2\bar{Y} \sum_{i=1}^n \hat{u}_i}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

となりますから、

$$\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \neq 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

を意味します。よって、決定係数として、左辺か右辺のどちらを用いるべきかわかりません。どちらが正しい当てはまりの尺度とも言えませんから、定数項がないモデルでは、決定係数の解釈には注意が必要です。

練習問題 9

最小 2 乗法は、残差 2 乗和である

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (Y_i - \tilde{\alpha} - \tilde{\beta}_1 X_{1i} - \tilde{\beta}_2 X_{2i})^2$$

を最小にするように、 $\tilde{\alpha}$ 、 $\tilde{\beta}_1$ 、 $\tilde{\beta}_2$ を選びます。まず、 $\tilde{\alpha}$ で微分して 0 と置くと

$$-2 \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

となり、両辺を -2 で割ると

$$\sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

となります。上式を満たす $\hat{\alpha}$ 、 $\hat{\beta}_1$ 、 $\hat{\beta}_2$ は最小 2 乗推定量なのでハットを付けています。この式を展開すると、

$$\sum_{i=1}^n Y_i - n\hat{\alpha} - \hat{\beta}_1 \sum_{i=1}^n X_{1i} - \hat{\beta}_2 \sum_{i=1}^n X_{2i} = 0$$

となりますから、両辺を n で割って、 $\hat{\alpha}$ について解くと、

$$\hat{\alpha} = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

となります。

次に、残差 2 乗和を $\hat{\beta}_1$ で微分して 0 と置くと、

$$-2 \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) X_{1i} = 0$$

となり、さらに両辺を -2 で割ると、

$$\sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) X_{1i} = 0$$

となります。ここで $\hat{\alpha}$ の式を上式に代入すると、

$$\begin{aligned} & \sum_{i=1}^n (Y_i - (\bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2) - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) X_{1i} \\ &= \sum_{i=1}^n ((Y_i - \bar{Y}) - \hat{\beta}_1 (X_{1i} - \bar{X}_1) - \hat{\beta}_2 (X_{2i} - \bar{X}_2)) X_{1i} \\ &= \sum_{i=1}^n (Y_i - \bar{Y}) X_{1i} - \hat{\beta}_1 \sum_{i=1}^n (X_{1i} - \bar{X}_1) X_{1i} - \hat{\beta}_2 \sum_{i=1}^n (X_{2i} - \bar{X}_2) X_{1i} = 0 \end{aligned}$$

となります。ここで偏差の和は 0 ですから³、上式は

$$\sum_{i=1}^n (Y_i - \bar{Y})(X_{1i} - \bar{X}_1) - \hat{\beta}_1 \sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1) - \hat{\beta}_2 \sum_{i=1}^n (X_{2i} - \bar{X}_2)(X_{1i} - \bar{X}_1) = 0$$

と書き換えられます。ここで左辺の第3項は、標本共分散が 0 と仮定していますから 0 です。よって、上式は

$$\sum_{i=1}^n (Y_i - \bar{Y})(X_{1i} - \bar{X}_1) - \hat{\beta}_1 \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 = 0$$

となり、これを $\hat{\beta}_1$ について解けば

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1)(Y_i - \bar{Y})}{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2}$$

が得られます。これは Y_i を X_{1i} だけで回帰したときの OLS 推定量の式に他なりません。つまり、 X_1 と X_2 の標本共分散が 0 であれば、 X_2 を無視して X_1 だけで回帰しても、 X_2 を含めて回帰しても、同じ結果が得られるのです。

12章では、含めるべき変数を考慮しないとき、推定にバイアスが生じるとしました。しかし、この問題から分かることは、説明変数間に相関がなければ、

³偏差の和は 0 ($\sum_{i=1}^n (X_{1i} - \bar{X}_1) = 0$, $\sum_{i=1}^n (X_{2i} - \bar{X}_2) = 0$) から、以下が成立することに注意してください。

$$\textcircled{1} \quad \sum_{i=1}^n (X_{1i} - \bar{X}_1) X_{1i} = \sum_{i=1}^n (X_{1i} - \bar{X}_1) X_{1i} - \sum_{i=1}^n (X_{1i} - \bar{X}_1) \bar{X}_1 = \sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1)$$

$$\textcircled{2} \quad \sum_{i=1}^n (X_{2i} - \bar{X}_2) X_{1i} = \sum_{i=1}^n (X_{2i} - \bar{X}_2) X_{1i} - \sum_{i=1}^n (X_{2i} - \bar{X}_2) \bar{X}_1 = \sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)$$

式展開では、 $\sum_{i=1}^n (X_{1i} - \bar{X}_1) \bar{X}_1 = \bar{X}_1 \sum_{i=1}^n (X_{1i} - \bar{X}_1) = \bar{X}_1 \times 0 = 0$ 、 $\sum_{i=1}^n (X_{2i} - \bar{X}_2) \bar{X}_1 = \bar{X}_1 \sum_{i=1}^n (X_{2i} - \bar{X}_2) = 0$ を用いた。

それらの変数を除いて推定しても何の問題もないことが分かります。換言すれば、説明変数間の相関があるときだけ、それらの変数を除いて推定すると問題が生じるのです。この結果は、説明変数が任意の k 個あっても成立します。

練習問題 8 の追加問題

(4) $\hat{\beta}_r$ の分散を求めよ。

これは $\hat{\beta}_r$ の確率的表現と期待値を用いることで導出できる。具体的には、

$$\begin{aligned} V(\hat{\beta}_r) &= E[(\hat{\beta}_r - E[\hat{\beta}_r])^2] \\ &= E\left[\left(\beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} + \frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2} - \left(\beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2}\right)\right)^2\right] \\ &= E\left[\left(\frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2}\right)^2\right] = \frac{\sum_{i=1}^n X_i^2 E[u_i^2]}{(\sum_{i=1}^n X_i^2)^2} = \frac{\sigma^2 \sum_{i=1}^n X_i^2}{(\sum_{i=1}^n X_i^2)^2} = \frac{\sigma^2}{\sum_{i=1}^n X_i^2} \end{aligned}$$

式展開では、 $E[u_i^2] = \sigma^2$ 、 i と j が異なると $E[u_i u_j] = 0$ となることを用いた。

(5) $\hat{\beta}_r$ と通常の OLS 推定量 $\hat{\beta}$ では、どちらの方が分散は大きいのか。

ここで $\hat{\beta}$ の分散は

$$V(\hat{\beta}) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

であり、 $\alpha=0$ という制約を課して推定した OLS 推定量 $\hat{\beta}_r$ の分散は

$$V(\hat{\beta}_r) = \frac{\sigma^2}{\sum_{i=1}^n X_i^2}$$

となりました。2 章練習問題 12(2)で確認した通り、 $\sum_{i=1}^n (x_i - a)^2$ を最小にする a は、 x の平均でした。このため、 $\sum (X_i - \bar{X})^2$ よりも、 $a=0$ とした $\sum_{i=1}^n X_i^2$ はより大きくなります。

$$\sum (X_i - \bar{X})^2 \leq \sum_{i=1}^n X_i^2$$

逆数で考えると

$$\frac{1}{\sum_{i=1}^n X_i^2} \leq \frac{1}{\sum (X_i - \bar{X})^2}$$

となりますから、両辺に σ^2 を掛けることで

$$V(\hat{\beta}_r) = \frac{\sigma^2}{\sum_{i=1}^n X_i^2} \leq \frac{\sigma^2}{\sum (X_i - \bar{X})^2} = V(\hat{\beta})$$

が得られます。 $\alpha=0$ という制約を課すことで、推定するパラメータが減り、パラメータは安定して推定されていることがわかります。

(5) $\hat{\beta}$ と $\hat{\beta}_r$ を比較してどちらが望ましい推定量といえるか。バイアスと分散という言葉を用いて論じなさい。

これまで確認した通り、 $\hat{\beta}_r$ は不偏推定量ではありませんが（バイアスがある）、分散の小さい推定量となっています。これに対し、 $\hat{\beta}$ は不偏推定量ではありますが（バイアスがない）、分散はより大きくなっています。では、どちらが望ましい推定量なのでしょうか。

推定量の良さを測る指標として、**平均 2 乗誤差**（Mean Squared Error、略して **MSE**）があります。

MSE を説明するため、あるパラメータ θ を推定する推定量 $\hat{\theta}$ を考えましょう。このとき、**MSE** は、推定量 $\hat{\theta}$ と真の値 θ との乖離の 2 乗の期待値として定義されます。

$$E[(\hat{\theta} - \theta)^2]$$

MSE が小さいほど、その推定量は良いと判断されます。ちなみに、 $E[\hat{\theta}]$ は真の値 θ と一致していない可能性があるので、**MSE** は $\hat{\theta}$ の分散ではないことに注意してください。

MSE の意味を考えてみましょう。**MSE** は、

$$\begin{aligned} E[(\hat{\theta} - \theta)^2] &= E\left[\left((\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta)\right)^2\right] \\ &= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + (E[\hat{\theta}] - \theta)^2 + 2(E[\hat{\theta}] - \theta)E[(\hat{\theta} - E[\hat{\theta}])] \\ &= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + (E[\hat{\theta}] - \theta)^2 \end{aligned}$$

と展開できます。式展開では、 $(E[\hat{\theta}] - \theta)$ は固定した値なので期待値の外に出せること、また $E[(\hat{\theta} - E[\hat{\theta}])] = E[\hat{\theta}] - E[\hat{\theta}] = 0$ であること、を用いました。ここで右辺第 1 項目は、推定量の分散ですから、

$$V(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2]$$

と表記します。右辺第 2 項目は $E[\hat{\theta}] - \theta$ が、推定量のバイアスを表している

ため、

$$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta$$

と表記しましょう。

以上から、MSE は

$$E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + Bias(\hat{\theta})^2$$

となります。つまり、MSE は、推定量の分散とバイアスに依存しています。推定量の分散はばらつきを表しており、推定量のバイアスは分布の中心が θ からずれている程度を表します。MSE の小さい良い推定量は、推定量の分散が小さく、バイアスも小さいことがわかります。

初級の計量経済学では、不偏推定量だけを扱いますので（バイアスは 0 となる）、推定量の分散が小さいことが重視されていました。しかし、上級の計量経済学では、たとえバイアスがあっても、そのバイアスが小さく、推定量の分散も小さいのであれば、良い推定量と判断されます。

たとえば、以下の図では、2 種類の推定量を示しています。(a)では不偏性は満たしていますが、分散は大きくなっています。(b)では少しバイアスがありますが、分散は小さくなっています。MSE でみると、(b)のほうが小さくなるでしょうから、(b)がより望ましい推定量といえます。

(a) バイアスなし、分散大きい

(b) バイアスあり、分散小さい



以上を踏まえて、 $\hat{\beta}$ と $\hat{\beta}_r$ のどちらがより望ましい推定量かを考えてみましょう。通常の OLS 推定量 $\hat{\beta}$ の MSE は

$$\begin{aligned} E[(\hat{\beta} - \beta)^2] &= V(\hat{\beta}) + Bias(\hat{\beta})^2 \\ &= \frac{\sigma^2}{\sum(X_i - \bar{X})^2} \end{aligned}$$

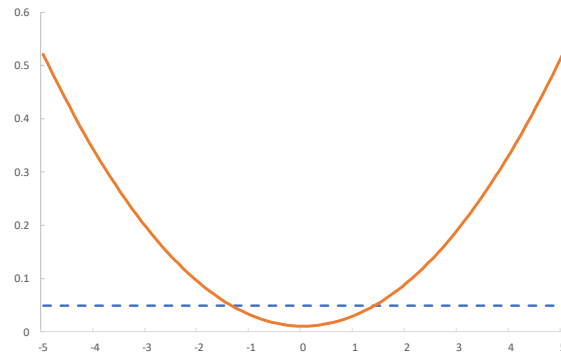
となります。ここでバイアスはないので、2 項目は 0 となっています。これ

に対し、 $\hat{\beta}_r$ の MSE は

$$\begin{aligned} E[(\hat{\beta}_r - \beta)^2] &= V(\hat{\beta}_r) + \text{Bias}(\hat{\beta}_r)^2 \\ &= \frac{\sigma^2}{\sum_{i=1}^n X_i^2} + \left(\frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} \right)^2 \end{aligned}$$

となります。ここで $\alpha = 0$ なら、 $\hat{\beta}_r$ の MSE の方が小さくなります。しかし、 α が 0 から乖離するほど、 $\text{Bias}(\hat{\beta}_r)^2$ は大きくなるため、 $\hat{\beta}_r$ の MSE は大きくなってしまいます。したがって、 α が 0 に十分に近いときは、 $\hat{\beta}_r$ の方が MSE は小さくなりますが、 α が 0 から十分に乖離すると $\hat{\beta}$ の方が望ましくなります。

どの程度、 α が 0 から乖離したら $\hat{\beta}$ の方が望ましいのでしょうか。これは X や σ^2 の値に応じて変わります。ここでは $n=10$ で $X_1 = 1, X_2 = 2, \dots, X_{10} = 10$ 、また、 $\sigma^2 = 4$ としてみましょう。以下の図では、横軸を α とし、縦軸に MSE を測っています。実線が $\hat{\beta}_r$ の MSE であり、点線が $\hat{\beta}$ の MSE となります。 $\hat{\beta}$ の MSE は、 α に依存しませんから、一定の値をとっています。これに対し、 $\hat{\beta}_r$ の MSE は、 α が 0 から乖離するにつれて、大きくなっていることがわかります。



実証分析をするとき、 α の値が事前に 0 である、もしくは 0 に非常に近い値であるとわかっていることはあまりありません。このため、 $\alpha = 0$ という制約を課してしまうことで、バイアスを発生させるようなリスクをとることは無いと思います。また、 $\alpha = 0$ という制約を課してしまうと、決定係数を用いることができないという欠点もあります。加えて、データ蓄積が急速に進んでいる現在、通常の OLS でも推定量の分散を小さく推定できるわけですから、リスクをとって $\alpha = 0$ という制約を課すことは避けた方がよいと思われます。