# 練習問題の解答

※解答の数値は有効桁のとり方や，分布表の精度によって計算誤差が生じることに注意されたい．

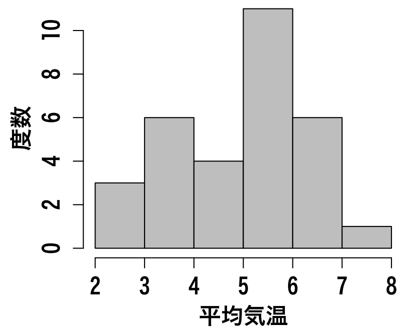
## 第 1 章

### 問 1.1

(1) 平均気温：量的変数，平均湿度：量的変数，日照時間：量的変数，風向き：質的変数．

(2) 相対度数の総計が 1.00 でないのは丸め誤差のためである．ヒストグラムの形よりベル型または少し左に裾が長い形になっていることがわかる．

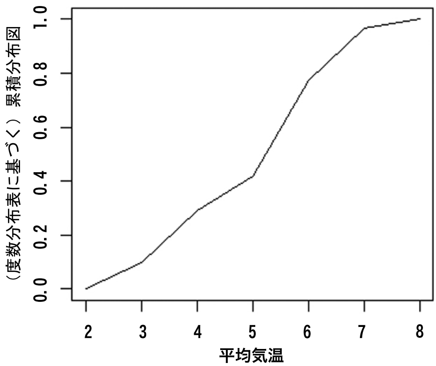
| 平均気温(℃) | 度数 | 相対度数 |
|---|---|---|
| 2.1〜3.0 | 3 | 0.10 |
| 3.1〜4.0 | 6 | 0.19 |
| 4.1〜5.0 | 4 | 0.13 |
| 5.1〜6.0 | 11 | 0.35 |
| 6.1〜7.0 | 6 | 0.19 |
| 7.1〜8.0 | 1 | 0.03 |
| 総計 | 31 | 0.99 |

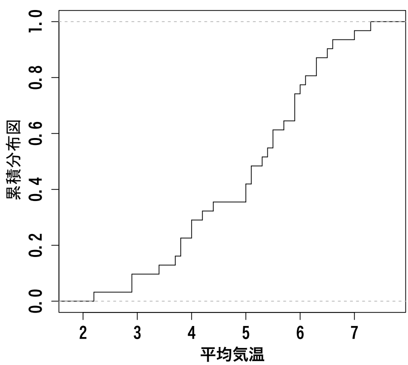



(3) 式 (1.3.1)，(1.3.2)，(1.3.3) を用いて計算する．平均気温：平均 $=$ 5.05 (℃)，分散 $=$ 1.62，標準偏差 $=$ 1.27 (℃)，平均湿度：平均 $=$ 36.2 (%)，分散 $=$ 69.5，標準偏差 $=$ 8.3 (%)，日照時間：平均 $=$ 7.87 (時間)，分散 $=$ 3.65，標準偏差 $=$ 1.91 (時間)．

(4)

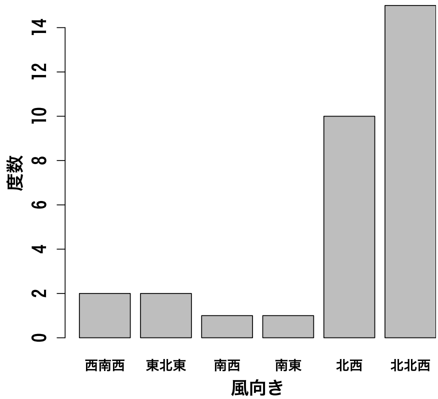
| 平均気温(℃) | 相対度数 | 累積相対度数 |
|---|---|---|
| 2.1〜3.0 | 0.10 | 0.10 |
| 3.1〜4.0 | 0.19 | 0.29 |
| 4.1〜5.0 | 0.13 | 0.42 |
| 5.1〜6.0 | 0.35 | 0.77 |
| 6.1〜7.0 | 0.19 | 0.96 |
| 7.1〜8.0 | 0.03 | 0.99 |
| 総計 | 0.99 | |



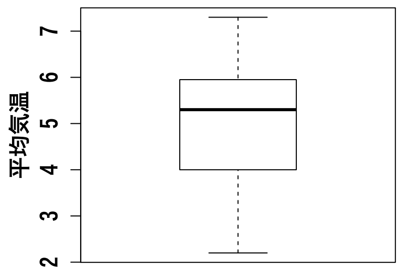



(5)

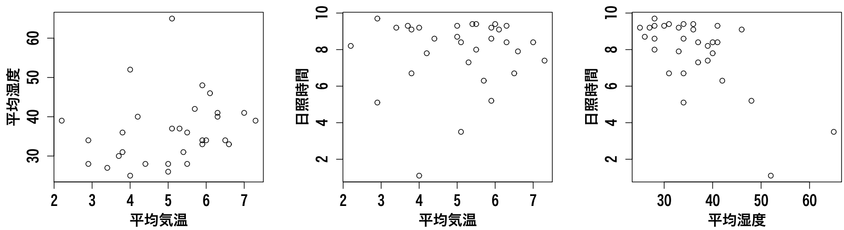
| 方角 | 度数 | 相対度数 |
|---|---|---|
| 西北西 | 2 | 0.06 |
| 東北東 | 2 | 0.06 |
| 南西 | 1 | 0.03 |
| 南東 | 1 | 0.03 |
| 北西 | 10 | 0.32 |
| 北北西 | 15 | 0.48 |
| 総計 | 31 | 0.98 |



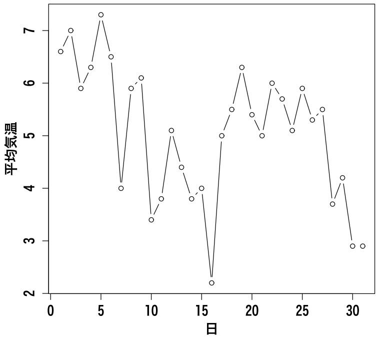

(6) 最小値 = 2.2，第 1 四分位数 = 4.0，第 2 四分位数 = 5.3，第 3 四分位数 = 6.00（または 5.95），最大値 = 7.3.




(7) 平均気温と平均湿度，平均気温と日照時間，平均湿度と日照時間の共分散はそれぞれ，2.41，0.16，$-10.71$ である．相関係数は共分散を両方の標準偏差で割ることより求められ，それぞれの相関係数は 0.23，0.07，$-0.67$ となる．この結果より，平均湿度と日照時間には比較的強い負の相関がみられる．




(8)




※四分位数の求め方には，大きく (A) 観測値を大きさの順に並べて求める方法，(B) 累積分布図に基づいて求める方法，の 2 種類がある．

(A) の中にもいくつかの方法があるが，p. 24 のコラムに解説されている Tukey の方法のアルゴリズムがわかりやすい．Tukey の方法では両側からの順位が出合うまでのカウント数を深度 depth と定義し，中央値 median の深度を depth(M) $=(n+1)/2=16$（設問では $n=31$）により求める．すなわち，中央値は順位 16 番目の観測値となる．次に全体を下半分 1〜16 番目と上半分 16〜31 番目の 2 群に分けて，それぞれ 16 個の観測値に対してその中での中心の位置（hinge，ちょうつがい）を求める．16 個の観測値の中心 hinge の深度は depth(H) $=([\text{depth(M)}]+1)/2=8.5$ となり，下半分では 8 番目と 9 番目，上半分では 23 番目と 24 番目の観測値の平均をそれぞれ第 1 四分位数，第 3 四分位数とする．

(B) の方法の場合，小問 (4) の度数分布表に基づく累積分布図を描いて $y$ 座標の $y=0.25, 0.50, 0.75$ に対する $x$ 座標の値を読み取り第 1，第 2，第 3 四分位数を求める．また，小問 (4) のもうひとつの累積分布図を用いてもよい．

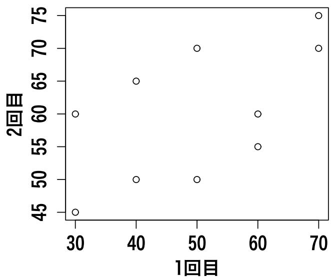
**問 1.2**

(1) 1 回目の平均：50.0（点）標準偏差：14.1（点），2 回目の平均：60.0（点）標準偏差：9.5（点）.

(2) C さんの 1 回目と 2 回目の得点はそれぞれ 40 点，50 点である．1 回目と 2 回目の標準化得点はそれぞれ $(40-50.0)/14.1 = -0.71$, $(50-60.0)/9.5 = -1.05$ となる.

(3) 1 回目と 2 回目の試験の変動係数 (=標準偏差/平均) は，それぞれ 0.28 (28%)，0.16（16%）である.

(4) 1 回目と 2 回目の共分散は 80.0 であり，相関係数 $= 80.0/(14.1 \times 9.5) = 0.60$ となる.




(5) 回帰直線の傾きは，共分散を 1 回目の試験の分散で割ることで求められ，$80.0/14.1^2 = 0.4$ となる．または，相関係数 × (2 回目の試験の標準偏差/1 回目の試験の標準偏差) で求められ，$0.6 \times (9.5/14.1) = 0.4$ となる．回帰直線は（1 回目の試験の平均，2 回目の試験の平均）を通るので，$y - 60.0 = 0.4(x - 50.0)$ を変形することで求められ，$y = 40 + 0.4x$ となる.

(6) 決定係数は (総平方和 − 残差平方和)/総平方和 で求められる．相関係数の 2 乗と等しくなることを用いて，$0.6^2 = 0.36$ となる.

**問 1.3**

(1) 軌道長半径の平均 = 8.45 (AU)，分散 = 102.85，標準偏差 = 10.14 (AU)，公転周期の平均 = 36.73 (年)，分散 = 3054.09，標準偏差 = 55.26 (年)，および，軌道長半径と公転周期の共分散 = 553.79 である（分散，共分散の計算はデータの大きさ $n$ で割っている)．これらより，回帰直線の傾きは $553.79/102.85 = 5.38$ であり回帰直線は $y - 36.73 = 5.38(x - 8.45)$ を変形することで求められ，$y = -8.73 + 5.38x$ となる（統計ソフトウェアで厳密に計算すれば，$y = -8.79 + 5.38x$

となる）．相関係数は 0.988 で，決定係数は 0.98 である．

(2)　(軌道長半径)$^2$の平均 $= 174.31$（$\mathrm{AU}^2$），分散 $= 89892.93$，標準偏差 $= 299.82$（$\mathrm{AU}^2$），および，公転周期との共分散 $= 16465.69$ である．これより，回帰直線の傾きは，$16465.69/89892.93 = 0.18$ であり，回帰直線は $y - 36.73 = 0.18(x - 174.31)$ を変形することで求められ，$y = 5.35 + 0.18x$ となる（統計ソフトウェアで厳密に計算すれば，$y = 4.80 + 0.18x$ となる）．相関係数は 0.994 で，決定係数は 0.99 である．

## 第 2 章

### 問 2.1

(1)　$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.5 + 0.7 - 0.9 = 0.3.$

(2)　$P(B \mid A) = P(A \cap B)/P(A) = 0.3/0.5 = 0.6.$

### 問 2.2

(1)　$0.080 \times 0.100 \times 0.01 = 0.00008.$

(2)　$0.080 \times 0.100/0.0368 = 0.2174.$（p. 64 の解答例参照）

(3)　$1 - 0.0368 = 0.9632.$

(4)　$1 - 0.0368 \times 0.01 = 0.99963.$

### 問 2.3

$x$ は二項分布 $B(n, p)$ に従い，その平均は $np$，分散は $np(1-p)$ である．よって，

$$\text{平均} = E[x/n] = E[x]/n = np/n = p$$
$$\text{分散} = V[x/n] = V[x]/n^2 = np(1-p)/n^2 = p(1-p)/n$$
$$\text{標準偏差} = \sqrt{p(1-p)/n}$$

である．

### 問 2.4

式 (2.7.9) と $e^{-2.1} = 0.12246$ より，交通事故数が 0 件，1 件，2 件，3 件，4 件以上の発生確率は，順に 0.122，0.257，0.270，0.189，0.161 である．

### 問 2.5

式 (2.8.5) により標準正規分布に従う確率変数に変換し，付表 1（標準正規分布の上側確率）を利用し求める．

(1) $P(60 < X \leq 70) = P(1 < Z \leq 2) = 0.1587 - 0.0228 = 0.1359.$

(2) 標本サイズ 100 の標本平均 $\bar{X}$ は $N(50, 1)$ に従う．$P(52 < \bar{X}) = P(2 < Z) = 0.0228.$

(3) 標本サイズ $n$ の標本平均 $\bar{X}$ は $N(50, 100/n)$ に従う．$P(52 < \bar{X}) = P(2\sqrt{n/100} < Z) = 0.05$ より，$2\sqrt{n/100} = 1.645$ を解くと $n \fallingdotseq 68$ となる．

**問 2.6**

体重が平均 $= 55.6$ (kg)，標準偏差 $= 7.2$ (kg) に従うことから，10 人の総体重 $S$ は平均 $= 556$ (kg)，標準偏差 $= 7.2 \times \sqrt{10} = 22.77$ (kg) に従う．$P(600 < S) = P((600 - 556)/22.77 < Z) = P(1.93 < Z) = 0.027$ となる．

## 第 3 章

**問 3.1**

(1) (A)：$62.2 \pm 1.960 \times 11.0/\sqrt{32}$ より，[58.4, 66.0] となる．(B)：$71.4 \pm 1.960 \times 10.8/\sqrt{35}$ より，[67.8, 75.0] となる．

(2) (A)：$62.2 \pm 2.040 \times 11.0/\sqrt{32}$ より，[58.2, 66.2] となる．(B)：$71.4 \pm 2.032 \times 10.8/\sqrt{35}$ より，[67.7, 75.1] となる．ここで，2.040 と 2.032 はそれぞれ自由度 31 と 34 の $t$ 分布の上側 2.5% 点である．

※自由度 31 と 34 の $t$ 分布の上側 2.5% の値は付表 2 にはないので補間による近似値を用いてもよい．

**問 3.2**

(1) 標本サイズが 2500 と大きいので，標本平均 $\overline{X}$ の分布は正規分布で近似できる．したがって，$32.8 \pm 1.960 \times 29.5/\sqrt{2500}$ より，[31.6, 34.0] となる．

(2) 消費支出のような強い歪みを持った母集団分布からの標本平均の分布の場合，調査世帯数が $n = 25$ 程度では，正規分布には十分近いとは言えない．したがって，正規分布を利用した信頼区間や自由度 $25 - 1 = 24$ の $t$ 分布を用いた信頼区間は参考程度の意味しか持たない．たとえば，$t$ 分布を用いた信頼区間は $32.8 \pm 2.064 \times 29.5/\sqrt{25}$ より，[20.6, 45.0] となる．

**問 3.3**

(1) $0.6 \pm 1.960 \times \sqrt{0.6 \times 0.4/2500}$ より，[0.581, 0.619] となる．

(2) 実際は非復元抽出を用いているとしても，母集団が $N = 100000$ と大きく抽出率 $f = n/N$ が小さいため，有限母集団修正の係数 $fpc = (N-n)/(N-1) \doteqdot 1 - f = 0.975$ は 1 に近いから，結論はほとんど変わらない．実際，係数 $fpc$ を用いて求めた信頼区間は [0.581, 0.619] となり，有限母集団修正を行わない方法と 3 ケタまで一致している．

(3) $0.6 \pm 1.960 \times \sqrt{(5000-2500)/(5000-1)} \times \sqrt{0.6 \times 0.4/2500}$ より，[0.586, 0.614] となる．

**問 3.4**

(1) プールした分散は $(17 \times 29.70 + 17 \times 34.12)/(17+17) = 31.91$ である．$(171.00 - 167.94) \pm 2.032 \times \sqrt{2 \times 31.91/18}$ より，$[-0.77, 6.88]$ となる．信頼区間に 0 が含まれている．$t$ 分布の自由度は 34 である．

(2) p. 85 の 1 行目の式で $a = 1, b = -1, c = 0$ と置いた式を用いると，不偏分散と不偏共分散より, 対応のある 2 標本の差の分散は $34.12+29.70-2\times 24.71 = 14.4$ である．$(171.00 - 167.94) \pm 2.110 \times \sqrt{14.4/18}$ より，[1.17, 4.94] となる．$t$ 分布の自由度は 17 である．(1) の結果と比較すると，(2) の信頼区間が狭く，また，信頼区間には 0 が含まれていないことがわかる．(1) と (2) の結果が異なるのは，親と子の身長に遺伝の影響があることが関係していると考えられる．

※差の平均を実際のデータより求め，3.056 とし計算をしている．

※例 11（p. 125）によると，この調査は，背の高い父親から背の高い息子が生まれやすいといった遺伝の影響が考えられるため，その影響を除いて世代間の身長の差を分析する目的で，男子大学生とその父親をペアとしてサンプリングしている．この場合，父親の標本と息子の標本は「独立な 2 標本」（小問 (1)）ではなく，「対応のある 2 標本」（小問 (2)）であり，(1) の「独立な 2 標本」とみなして分析することは間違った分析と言える．(1) と (2) の分析結果の違いには遺伝の影響が大きく関わっていると考えられるが，遺伝の影響についての，より直接的分析としては，親子の身長の相関係数の分析，あるいは親と子の身長をそれぞれ説明変数と応答変数とする線形回帰分析を利用するほうが適切である．

**問 3.5**

(1) $\mu_1$ の信頼区間は $144.8 \pm 1.976 \times \sqrt{434.537/150}$ より，[141.4, 148.2] となる．

$\mu_2$ の信頼区間は $155.6 \pm 1.976 \times \sqrt{480.10/150}$ より，[152.1, 159.1] となる．$t$ 分布の自由度は 149 である．

(2) 自由度 149 の $\chi^2$ 分布の下側および上側 2.5% 点は 117.1, 184.7 である．これより，$\sigma_1^2$ の信頼区間は，$149 \times 434.537/184.7 = 350.5$ と $149 \times 434.537/117.1 = 552.9$ の間，つまり [350.5, 552.9] となる．$\sigma_2^2$ の信頼区間は，$149 \times 480.10/184.7 = 387.3$ と $149 \times 480.10/117.1 = 610.9$ の間，つまり [387.3, 610.9] となる．

(3) プールした分散は $(434.537 + 480.10)/2 = 457.319$ である．$\delta$ の信頼区間は $(144.8 - 155.6) \pm 1.968 \times \sqrt{2 \times 457.319/150}$ より，[−15.66, −5.94] となる．$t$ 分布の自由度は 298 である．

(4) 対応のある標本における $\delta$ の信頼区間は $-10.8 \pm 2.776 \times 3.91/\sqrt{5}$ より，[−15.65, −5.95] となる．$t$ 分布の自由度は 4 である．

## 第 4 章

### 問 4.1

20 回中，色が的中する回数 $x$ は二項分布 $B(20, p)$ に従う．帰無仮説 $H_0$ は「超能力を持っていない」こと，つまり $H_0 : p = 1/2$，対立仮説は $H_1 : p \neq 1/2$ となる．$H_0$ の下で $x \sim N(10, 5)$ と近似でき，有意水準 5% の検定の棄却域は $|x - 10|/\sqrt{5} \geq 1.96$ となる．観測値から $|15 - 10|/\sqrt{5} = 2.236$ となり，帰無仮説は有意水準 5% で棄却される．

### 問 4.2

本文より，$n = 20$，観測値の平均 $= 909.0$，標準偏差 $= 104.9$ である．帰無仮説 $H_0 : \mu = \mu_0$ の下で $t$ 値は自由度 19 の $t$ 分布に従い，有意水準 5% の検定の棄却域は $|t| \geq t_{0.025}(19) = 2.093$ となる．第 1 の帰無仮説 $H_0 : \mu = 990$ の下では，$t = \sqrt{20}(909.0 - 990)/104.9 = -3.45$ であり，帰無仮説は有意水準 5% で棄却される．第 2 の帰無仮説 $H_0 : \mu = 792$ の下では，$t = \sqrt{20}(909.0 - 792)/104.9 = 4.99$ であり，帰無仮説は有意水準 5% で棄却される．

### 問 4.3

(1) 帰無仮説 $H_0 : \mu = 500$，対立仮説 $H_1 : \mu < 500$．棄却域は $z = \sqrt{40}(\bar{x} - 500)/100 \leq -1.645$ となる．$z = -0.9486$ であり，帰無仮説は有意水準 5% で棄却されない．英語力が低下したとは判断できない．

(2) 選んだクラスが新入生全体から無作為に選ばれていない，帰国子女のクラスや

付属高校からの進学クラスを選んだりすると母集団を代表する標本とは言えない．標本は母集団から無作為に抽出されることが，統計的分析の正当性を保証する一つの方法である．単純無作為抽出の方法以外にも，各クラスから男女 2 名ずつを抽出するなど，さまざまな抽出方法が客観的な判断を可能とする目的で導入される．

**問 4.4**

(1) 帰無仮説 $H_0 : \mu = 200$，対立仮説 $H_1 : \mu \neq 200$．棄却域は $|z| > 1.96$ となる．$z = \sqrt{16}(207 - 200)/10 = 2.80$ であり，帰無仮説は有意水準 5% で棄却される．

(2) 信頼区間は $207 \pm 1.96 \times 10/\sqrt{16}$ より，[202.1, 211.9] となる．95% 信頼区間に 200 が含まれないことから，(1) の結果と同じであることがわかる．

(3) 帰無仮説 $H_0 : \mu = 200$，対立仮説 $H_1 : \mu \neq 200$．棄却域は $|t| > t_{0.025}(15) = 2.131$ となる．$t = \sqrt{16}(207 - 200)/10 = 2.80$ であり，帰無仮説は有意水準 5% で棄却される．95% 信頼区間は $207 \pm 2.131 \times 10/\sqrt{16}$ より，[201.7, 212.3] となる．(2) の信頼区間より広いが 200 は含まれない．

**問 4.5**

(1) 期待値 $= \mu_1 - \mu_2 = 0$，分散 $= \sigma^2(1/n_1 + 1/n_2)$．

(2) $d - 1.96\sigma\sqrt{1/n_1 + 1/n_2} \leq \delta \leq d + 1.96\sigma\sqrt{1/n_1 + 1/n_2}$.

(3) 帰無仮説 $H_0 : \delta = 0$，対立仮説 $H_1 : \delta \neq 0$．$z = \sqrt{10}(65.0 - 70.0)/6.32 = -2.50 < -1.96$ であり，帰無仮説は有意水準 5% で棄却される．

**問 4.6**

(1) 帰無仮説 $H_0$ の下で $\chi^2 = (n-1)\hat{\sigma}^2/\sigma^2$ が自由度 $n-1$ の $\chi^2$ 分布に従う．教材 $E_1$ では $\chi^2 = 149 \times 434.537/400 = 161.9$，教材 $E_2$ では $\chi^2 = 149 \times 480.10/400 = 178.8$ となる．自由度 149 の $\chi^2$ 分布の下側および上側 2.5% 点はそれぞれ 117.1，184.7 であることより，棄却域は 117.1 以下，または 184.7 以上である．これより，どちらも帰無仮説は有意水準 5% で棄却されない．

(2) プールした分散は $(434.537 + 480.10)/2 = 457.319$ である．$t = \sqrt{150}(144.8 - 155.6)/\sqrt{2 \times 457.319} = -4.37$ を棄却域 $|t| > t_{0.025}(298) = 1.97$ と比較する．帰無仮説は有意水準 5% で棄却される．

(3) $t = \sqrt{5}(-10.8)/3.91 = -6.18$ を棄却域 $|t| > t_{0.025}(4) = 2.78$ と比較する．帰

無仮説は有意水準 5% で棄却される．

※帰無仮説が両側検定において棄却されないということは，その信頼区間が 0 を含まないことと同等である．問 3.5(3)(4) を参照のこと．

## 第 5 章

### 問 5.1

回帰分析に関する統計解析ソフトウェアの出力結果については，本文中 (p. 179) にある説明を参照のこと．

### 問 5.2

(1) 平均年棒に対する回帰係数の $P$-値が 0.3107 であることから，平均年棒が勝率に関係しているとは考えにくい．決定係数や $F$-値に対する $P$-値から判断してもよいモデルとは言えない．

(2) $F$-値に対する $P$-値が 0.656 であることから，リーグによって平均年棒に差があるとは言えない．

### 問 5.3

(1) モデル 1：家賃 $= 38.16099 + 2.81737 \times$ 大きさ $- 0.52150 \times$ 徒歩 $- 1.16953 \times$ 築年数

モデル 2：家賃 $= 33.42282 + 2.83278 \times$ 大きさ $- 1.18052 \times$ 築年数

(2) 各説明変数に対する回帰係数の $P$-値，自由度調整済み決定係数，$F$-値に対する $P$-値などを比較してもあまり大きく違わないが，自由度調整済み決定係数からの判断では僅差でモデル 1 の方がよい．一方で，説明変数は少ない方がよいという考え方で判断するとモデル 2 でも問題ないと言える．

### 問 5.4

標本の相関係数 $r = 24.7/\sqrt{29.7 \times 34.1} = 0.776$ より，$z = \log((1 + 0.776)/(1 - 0.776))/2 = 1.035$ となる．相関係数 $\rho$ に対する $\zeta = \log((1+\rho)/(1-\rho))/2$ の信頼区間は $1.035 \pm 1.96/\sqrt{18-3}$ より，[0.529, 1.541] となる．元に戻すため tanh を施すことによって，相関係数 $\rho$ の 95% 信頼区間 [0.48, 0.91] を得る．

## 第 6 章

### 問 6.1

(1)　式 (2.7.9) と $e^{-2} = 0.13534$ より，交通事故死者数が 0 人，1 人，2 人，3 人，4 人以上の発生確率は，順に 0.135，0.271，0.271，0.180，0.143 である．

(2)　統計量の値は $(21-13.5)^2/13.5+(32-27.1)^2/27.1+\cdots+(6-14.3)^2/14.3 = 11.99$，自由度 4 の $\chi^2$ 分布の上側 5% 点は 9.49 であり，有意水準 5% で帰無仮説は棄却され，平均 2 のポアソン分布に従うとは言えない．

(3)　表から平均は $(0 \times 21 + 1 \times 32 + \cdots + 4 \times 6)/100 = 1.5$ と推定される．$e^{-1.5} = 0.22313$ より，交通事故死者数が 0 人，1 人，2 人，3 人，4 人以上の発生確率は，順に 0.223，0.335，0.251，0.126，0.066 である．

(4)　統計量の値は $(21-22.3)^2/22.3+(32-33.5)^2/33.5+\cdots+(6-6.6)^2/6.6 = 0.82$，自由度 3 の $\chi^2$ 分布の上側 5% 点は 7.81 であり，有意水準 5% で帰無仮説は棄却されないため，平均 1.5 のポアソン分布に従わないとは言えない．自由度については，p. 204 のコメント (2) を参照のこと．

### 問 6.2

(1)　1 行 1 列目から順に，$63 \times 92/120 = 48.3$, $63 \times 28/120 = 14.7$，$57 \times 92/120 = 43.7$，$57 \times 28/120 = 13.3$ である．

(2)　統計量の値は $(52-48.3)^2/48.3+(11-14.7)^2/14.7+\cdots+(17-13.3)^2/13.3 = 2.56$，自由度 1 の $\chi^2$ 分布の上側 5% 点は 3.84 であり，有意水準 5% で帰無仮説は棄却されない．これより，これらの薬に差があるとは言えない．

(3)　式 (4.4.7) の検定統計量を用いる．薬剤 A および薬剤 B の改善の割合は，それぞれ $52/63 = 0.825$, $40/57 = 0.702$ である．差がないとすると，改善の割合は $\hat{p}^* = 92/120 = 0.767$ であり，$\hat{p}^*(1-\hat{p}^*) = 0.767 \times 0.233$ となる．また，$\sqrt{1/63+1/57}$ を考慮し，割合の差の検定における検定統計量を求めると，$z = (0.825-0.702)/\sqrt{0.767 \times 0.233 \times (1/63+1/57)} = 1.60 < 1.96$ であり，有意水準 5% で帰無仮説は棄却されない．これより，これらの薬に差があるとは言えない．(2) と (3) の $P$-値は 0.110 と同じであり，同じ検定結果となる．