理解を確かめるための設問

問 1.1 方法論が必要となる例： 身近な事柄で方法の選択を余儀なくされる例を示し，考えられる選択肢を 3 つ示せ．

解答例： 健康を保持するために，毎週木曜夕方 2 時間を運動に費やすことにする．生活と職場環境を考えると選択肢は，(1) ジョギング，(2) テニス，(3) フィットネスクラブ，である．

注意：選択肢は，排他的，すなわち両方同時に採用することができないもの，でなければならない．

問 1.2 後智恵で恣意的な結論： 例 1.1 のデータにおいて，4 地域の頭髪水銀濃度が同程度である，という結論が導けるような客観的な代表値の定め方を考えよ．

解答例： 各地域で，順序統計量の中央の 14 個の測定値の平均を代表値にする．ただし取り除く個数が奇数の場合は，大きい値の方を 1 個多く除くことにする．そうすると代表値が，A_1 : 8.96，A_2 : 8.39，A_3 : 8.55，A_4 : 9.65 となり，4 地域の頭髪水銀濃度がほとんど同程度になる．

注意：結果をみてから自分が出したい結論に合う方法を探すと，それなりに見つかることがあるが，それは真実を反映しないことが多い．

問 1.3 ばらつきの原因： 体重が気になっているある人が，自分の体重を一日 1 回，7 日間測定を行ったところ，$\{60.2kg, 59.4kg, 62.0kg, 61.3kg, 58.8kg, 61.7kg, 60.6kg\}$ というデータが得られた．この値のばらつきの原因として考えられる因子を 3 つあげよ．その因子のうち，測定法を標準化することでばらつきを減らせるものがどれであるか説明せよ．

解答例： (1) 測定を食後に行ったか，食前に行ったか，(2) 測定を排泄前に行ったか，排泄後に行ったか，(3) 衣服の量，(4) 前夜の節制の有無，など．(1)〜(3) は標準化できるが，(4) はできないのが普通である．

問 1.4 統計量の特徴： $x_1 = 3$，$x_2 = 2$，$x_3 = 1$，$x_4 = 10$ というデータについて，算術平均（通常の平均），幾何平均，中央値を計算して，大小関係を比較し，データのどのような特徴を反映してこの大小関係が生じたかを説明せよ．

解答例： この数値例では，算術平均：4.00，幾何平均：2.78，中央値：2.50，で，中央値が最も小さく，算術平均が最も大きい．測定値の中に外れ値があるとき，算術平均はその影響を最も大きく受けて，代表値として望ましくない値となる．中央値は外れ値がどれくらい極端であってもその影響を受けない．算術平均と中央値が大きく異なるときは，外れ値の内容を丁寧に検討すべきである．

問 1.5 同じ統計量の異なる表現式： 次に示す等式を証明せよ．

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}x_i^2-n\bar{x}^2, \quad \bar{x}=\sum_{i=1}^{n}x_i/n \\
\sum_{i=1}^{n}(y_i-\bar{y})^2 &= \sum_{i=1}^{n}y_i^2-\frac{1}{n}(\sum_{j=1}^{n}y_j)^2, \quad \bar{y}=\sum_{i=1}^{n}y_i/n \\
\sum_{i=1}^{n}(y_i-a)^2 &= \sum_{i=1}^{n}(y_i^2-\bar{y})^2+n(\bar{y}-a)^2
\end{aligned}$$

解答例：

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)=\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2 \\
&= \sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2 \\
\sum_{i=1}^{n}(y_i-\bar{y})^2 &= \sum_{i=1}^{n}(y_i^2-2\bar{y}y_i+\bar{y}^2) \\
&= \sum_{i=1}^{n}y_i^2-2(\sum_{j=1}^{n}y_j/n)(\sum_{i=1}^{n}y_i)+n(\sum_{j=1}^{n}y_j/n)^2 \\
&= \sum_{i=1}^{n}y_i^2-\frac{1}{n}(\sum_{j=1}^{n}y_j)^2 \\
\sum_{i=1}^{n}(y_i-a)^2 &= \sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-a)^2 \\
&= \sum_{i=1}^{n}(y_i-\bar{y})^2-2(\bar{y}-a)\sum_{i=1}^{n}(y_i-\bar{y})+\sum_{i=1}^{n}(\bar{y}-a)^2 \\
&= \sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-a)^2
\end{aligned}$$

問 1.6 チェビシェフの不等式： n 個の数値 $x_1, x_2, \ldots, x_n$ の平均を $\bar{x}$，標準偏差を s と書くとき，任意の正数 a に対して,「区間 $(\bar{x}-as, \bar{x}+as)$ の外にある数値の個数は $(n-1)/a^2$ 以下である」という関係が成り立つ．この関係を “チェビシェフの不等式” (Chebyshev’s inequality) という．

(1) この不等式を証明せよ．

(2) ある試験で，100 人の生徒の得点の平均が 50 で標準偏差が 10 であるとき，80 点以上あるいは 20 点以下の生徒数が 11 人以下であることを示せ．ここで $s^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ である．

解答例： (1) 範囲外の測定値の個数を m とする．

$$(n-1)s^2=\sum_{i=1}^{n}(x_i-\bar{x})^2 \geq \sum_{\text{範囲外の測定値}}(as)^2=ma^2s^2 \quad \rightarrow \frac{n-1}{a^2}\geq m$$

(2) $a=3$，$n=100$ のとき，$(n-1)/a^2=11$ である．

注意 現在の高等学校の数学 I では，分散を $\sum_{i=1}^{n}(x_i-\bar{x})^2/n$ で定義している．この場合は，問題と解答の $n-1$ を n に置き換えることになる．

理解を確かめるための設問

問 2.1 確率変数の例： 身近な事柄で確率変数と想定できそうな例と，それが従うと考えられる確率分布を示せ．

解答例： 卓球のシングルスの試合で n 回サービスをするときのフォールトの数を Y とする．フォールト率を π とすると，Y は二項分布 $B(n,\pi)$ に従うと考えられる．

問 2.2 回転針： 平面上に円盤があり，周辺に 0 から 1 までの目盛りが等間隔に刻まれているとする．円盤の中心に軸を持ち，どの方向に対しても止まる確率が同じである回転可能な針があるとする．これを回転させて止まったときに針の指している目盛りを X とする．X は，確率密度が 0 と 1 の間で一様な連続分布（単位一様分布）に従う確率変数と考えられる．このとき，0 と 1 の間の任意の値 a に対して，$Pr\{X=a\}=0$ であることを，背理法で証明せよ．

（**ヒント**：背理法とは，命題が誤りであるとすると矛盾が起きることを示すことで，命題が正しいことを証明するやり方である．）

解答例： 任意の値 a に対して，$Pr\{X=a\}=\delta>0$ とすると，$\delta>1/n$ となる十分大きな正整数 n が存在する．$m>n$ を満たす正整数 m を用いて，区間 $(0,1)$ の間に m 個の点を取ると，どの点においても確率が δ だから，$Pr\{0<X<1\}>m\delta>m/n>1$ が成り立つことになる．これは確率の公理に矛盾するから $\delta=0$ でなければならない．

問 2.3 正規分布における標準化： X が正規分布 $N(\mu,\sigma^2)$ に従う確率変数であるとき，次式が成り立つことを示せ．

$$Pr\left\{\frac{X-\mu}{\sigma}<a\right\}=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}\ \exp\left(-\frac{u^2}{2}\right)\,du$$

解答例：

$$\begin{aligned}Pr\left\{\frac{X-\mu}{\sigma}<a\right\} &= Pr\{\ X<\mu+a\sigma\}\\ &= \int_{-\infty}^{\mu+a\sigma}\frac{1}{\sqrt{2\pi\sigma^2}}\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx\end{aligned}$$

積分変数を x から $u=(x-\mu)/\sigma$ に変換すると，積分上限が a になり，$du=dx/\sigma$ になるから与式が得られる．

問 2.4 カイ二乗分布： 確率密度関数が次式で定義される分布族をカイ二乗分布という．

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \exp\left(-\frac{x}{2}\right) x^{\nu/2-1}, \quad 0 < x < \infty, \quad 0 < \nu$$

ここで，Γ という記号で書かれているものは，“ガンマ関数” (gamma function) という名の特殊関数で，任意の正数 a に対して，$\Gamma(a) = \int_0^\infty \exp(-u)\, u^{a-1} du$ で定義される関数である．この関数 $f(x)$ が確率密度関数の性質を満たしていることを示せ．

注意：この式の中の ν という変数は，カイ二乗分布の母数で，“自由度” (degrees of freedom) と呼ばれている．

解答例： 式中の各関数は $x > 0, \nu > 0$ のときすべて正であるから $f(x) > 0$．$f(x)$ を $0 < x < \infty$ で積分したものが 1 になることは，積分変数を x から $u = x/2$ に変換することで，次のように示される．

$$\int_0^\infty \frac{1}{2^{\nu/2}} \exp\left(-\frac{x}{2}\right) x^{\nu/2-1} dx = \int_0^\infty \exp(-u)\, u^{\nu/2-1} du = \Gamma(\nu/2)$$

問 2.5 母集団の分布と標本の分布： ある診療所に登録されている糖尿病の患者は男が 40 人，女が 60 人であった．この患者からランダムに 10 人を選んだとき，男が丁度 4 人である確率を示せ．

解答例： 可能なすべての場合の数を分母，条件を満たす場合の数を分子，として比をとることで次の結果が得られる．$({}_{40}C_4 \times {}_{60}C_6)/{}_{100}C_{10} = 0.264$

注意：40% が男である集団からランダムに 10 人を選ぶと男が 4 人含まれると思いこむ人がいるが，その確率はそれほど大きくないことを数値例で示したものである．

問 2.6 偏差値： テストの得点などで，平均値が 50 点，標準偏差が 10 点となるように点数を基準化した値を偏差値という．テストの得点が正規分布しているときに偏差値が 60 以上の学生は全体の何%になるか，説明せよ．

解答例： 正規分布 $N(50, 10^2)$ に従う確率変数では，$U = (X-50)/10$ が標準正規分布 $N(0,1)$ に従うという事実を利用して，$Pr\{X > 60\}$ の確率を求めればよい．$Pr\{X > 60\} = Pr\{U > 1\} = 0.159$ であるから答は 15.9%である．

理解を確かめるための設問

問 3.1 モーメントの間の関係： 1 次元分布の場合，平均回りの 2 次のモーメント ν_2 と原点回りの 1, 2 次モーメント μ_1, μ_2 の間に $\nu_2 = \mu_2 - \mu_1^2$ という関係があることを示せ．

解答例： 当該確率分布に従う確率変数を Y とすると次のように式変形ができる．

$$\nu_2 = E\{(Y-\mu_1)^2\} = E\{Y^2 - 2Y\mu_1 + \mu_1^2\} = E\{Y^2\} - 2E\{Y\}\mu_1 + \mu_1^2 = \mu_2 - \mu_1^2$$

問 3.2 一様分布のモーメント： 確率密度関数が次式で与えられる連続分布を区間 (α,β) の一様分布という．

$$f(x) = \frac{1}{\beta-\alpha}, \quad \alpha < x < \beta, \quad -\infty < \alpha < \beta < \infty$$

この一様分布の平均と分散を計算せよ．

解答例：

$$\begin{aligned}
\mu &= \int_\alpha^\beta \frac{x}{\beta-\alpha}dx = \frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_\alpha^\beta = \frac{\beta+\alpha}{2} \\
\sigma^2 &= \int_\alpha^\beta \frac{x^2}{\beta-\alpha}dx - \mu^2 = \frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_\alpha^\beta - \mu^2 = \frac{\beta^3-\alpha^3}{3(\beta-\alpha)} - \left(\frac{\beta+\alpha}{2}\right) = \frac{(\beta-\alpha)^2}{12}
\end{aligned}$$

問 3.3 独立な確率変数に関する期待値： X と Y が独立な確率変数であるとき，$Cov\{X,Y\}=0$ であることを示せ．さらに，任意の関数 $s(X)$，$t(Y)$ について，$E\{s(X)t(Y)\} = E\{s(X)\}E\{t(Y)\}$ であることを示せ．

解答例： 簡単のために，X,Y の確率分布を連続型と仮定し，密度関数を $f(x), g(y)$ とおき，積分範囲の記載を省略し，後半を先に証明して，前半をその後に証明する．離散型の場合は積分を和にすればよい．

$$\begin{aligned}
E\{s(X)t(Y)\} &= \int\int s(x)t(y)f(x)g(y)dxdy \\
&= \left(\int s(x)f(x)dx\right)\left(\int t(y)g(y)dy\right) = E\{s(X)\}E\{t(Y)\} \\
Cov\{X,Y\} &= E\{(X-E\{X\})(Y-E\{Y\})\} = E\{X-E\{X\}\}E\{Y-E\{Y\}\} \\
&= (E\{X\}-E\{X\})(E\{Y\}-E\{Y\}) = 0
\end{aligned}$$

問 3.4 無相関と独立性の関係： X と Y の相関係数が 0 であるのに両者が独立でない例を示せ．

解答例： 2 次元の離散型確率変数 (X,Y) の確率関数 $f(x,y)$ を，$f(1,1)=1/4$，$f(0,-1)=1/2$，$f(-1,1)=1/4$ とする．それぞれの周辺確率関数 $f_1(x), f_2(y)$ は次のようになる．

$$f_1(1)=1/4, f_1(0)=1/2, f_1(-1)=1/4; f_2(1)=1/2, f_2(-1)=1/2$$

$f(x,y)=f_1(x)\times f_2(y)$ という関係が成り立たないので，X と Y は独立でない．

$E\{X\}=0, E\{Y\}=0$ であるから次式が成り立つ．

$$Cov\{X,Y\} = 1\times 1\times 1/4 + (-1)\times 1\times 1/4 = 0$$

共分散が 0 であるから，相関係数も 0 である．すなわち X と Y は独立でないのに相関係数は 0 である．

問 3.5 独立性： X_1, X_2, X_3 は互いに独立に同じ確率分布に従う確率変数であるとする．$Y_1 = X_1$，$Y_2 = X_2 - X_1$，$Y_3 = X_3 - X_1$，$Y_4 = X_1 + X_2$ とする．X_i $(i = 1, 2, 3)$ の標本空間が $\{1, 2, 3\}$ で，x の確率 $f(x)$ が，$f(1) = 1/4$，$f(2) = 1/2$，$f(3) = 1/4$ で与えられるとき，Y_1, Y_2, Y_3, Y_4 のどれとどれが独立か，検討せよ．

解答例： X_i の分散を σ^2 とする．

$$
\begin{aligned}
Cov\{Y_1, Y_2\} &= Cov\{X_1, X_2 - X_1\} = V\{X_1\} = \sigma^2 > 0 \\
Cov\{Y_1, Y_3\} &= Cov\{X_1, X_3 - X_1\} = V\{X_1\} = \sigma^2 > 0 \\
Cov\{Y_1, Y_4\} &= Cov\{X_1, X_2 + X_1\} = V\{X_1\} = \sigma^2 > 0 \\
Cov\{Y_2, Y_3\} &= Cov\{X_2 - X_1, X_3 - X_1\} = V\{X_1\} = \sigma^2 > 0 \\
Cov\{Y_2, Y_4\} &= Cov\{X_2 - X_1, X_2 + X_1\} = V\{X_2\} - V\{X_1\} = 0 \\
Cov\{Y_3, Y_4\} &= Cov\{X_3 - X_1, X_2 + X_1\} = V\{X_1\} = \sigma^2 > 0
\end{aligned}
$$

独立の可能性があるのは，Y_2 と Y_4 だけであるから，これについて同時確率 $p(y_2, y_4)$ を計算すると，以下の表になる．たとえば，$Pr\{Y_2 = 2\} = 1/16$, $Pr\{Y_4 = 2\} = 1/16$ であるのに，$Pr\{Y_2 = 2, Y_4 = 2\} = 0$ であるから，$Pr\{Y_2 = 2, Y_4 = 2\} \neq Pr\{Y_2 = 2\} \times Pr\{Y_4 = 2\}$ である．したがって Y_2 と Y_4 も独立ではない．

		X_2		
		1	2	3
X_1	1	$(Y_2, Y_4) = (0, 2)$ $p(0, 2) = 1/16$	$(Y_2, Y_4) = (1, 3)$ $p(1, 3) = 2/16$	$(Y_2, Y_4) = (2, 4)$ $p(2, 4) = 1/16$
	2	$(Y_2, Y_4) = (-1, 3)$ $p(-1, 3) = 2/16$	$(Y_2, Y_4) = (0, 4)$ $p(0, 4) = 4/16$	$(Y_2, Y_4) = (1, 5)$ $p(1, 5) = 2/16$
	3	$(Y_2, Y_4) = (-2, 4)$ $p(-2, 4) = 1/16$	$(Y_2, Y_4) = (-1, 5)$ $p(-1, 5) = 2/16$	$(Y_2, Y_4) = (0, 6)$ $p(0, 6) = 1/16$

問 3.6 統計量の期待値と分散： X をさいころの出る目の数としたとき，$Y = 4 + X$ は確率変数であるといってよいか．もしよければ，Y の期待値と分散を計算せよ．

解答例： Y の取り得る値の全体，すなわち標本空間は，$\{5, 6, 7, 8, 9, 10\}$ で，それぞれの値を取る確率はすべて $1/6$ であるから，確率変数である．

$$
\begin{aligned}
E\{Y\} &= (5 + 6 + 7 + 8 + 9 + 10)/6 = 7.5 \\
V\{Y\} &= [(5 - 7.5)^2 + (6 - 7.5)^2 + \cdots + (10 - 7.5)^2]/6 = 2.92
\end{aligned}
$$

問 3.7 統計量の確率分布： 確率変数 X の確率密度関数が，$0 \geq x$ のとき $f(x) = 0$，$0 < x$ のとき $f(x) = \exp(-x)$ であるとする．$X < 1$ ならば $Y = 0$，$X \geq 1$ ならば $Y = 1$ として定義される確率変数 Y の期待値を計算せよ．

解答例：

$$E\{Y\} = 1 \times Pr\{Y = 1\} = Pr\{X \geq 1\} = \int_1^\infty \exp(-x)dx = \exp(-1)$$

理解を確かめるための設問

問 4.1 平均二乗誤差の性質： 標本 $\boldsymbol{Y}$ が従う確率分布が，θ を母数とする分布族に含まれているとする．θ の 1 つの値 θ_0 に対しては，平均二乗誤差を 0 とする推定量が作れることを示せ．これに基づいて，すべての θ において他の任意の推定量より平均二乗誤差が小さい推定量，というものが作れないことを説明せよ．

解答例： $T = \theta_0$ を一般の θ の推定量とする．この推定量の平均二乗誤差は $(\theta_0 - \theta)^2$ であるから，$\theta = \theta_0$ のときこの値は 0 になる．θ_0 は任意であるから，平均二乗誤差をどのような θ に対しても最小にする推定量があるとすると，その推定量の平均二乗誤差は常に 0 でなければならない．これは θ の真の値が既知でない限り不可能である．

問 4.2 最良線形不偏推定量： $Y_1, Y_2, \ldots, Y_n$ が互いに独立に

$$E\{Y_i\} = \mu, V\{Y_i\} = \sigma^2; (i = 1, 2, \ldots, n)$$

の分布に従う確率変数であるとき，線形推定量 $T = c_1 Y_1 + c_2 Y_2 + \cdots + c_n Y_n$ が μ の不偏推定量であるための $c_1, c_2, \ldots, c_n$ が満たすべき条件を求めよ．また，$V\{T\}$ を最小にする $c_1, c_2, \ldots, c_n$ の値を求めよ．

（**ヒント：** $n = 3$ の場合について，(c_1, c_2, c_3) を 3 次元空間の座標と考え，幾何学的に解を求め，それを一般化して推定量の見当をつけ，次の不等式を使ってその妥当性を証明するのが 1 つの方法である.）

$$\sum_{i=1}^n c_i^2 \geq n\bar{c}^2, \quad \bar{c} = \sum_{i=1}^n c_i/n$$

解答例： 不偏性の条件は $c_1 + c_2 + \cdots + c_n = 1$ である．任意の線形推定量の分散は，$(c_1^2 + c_2^2 + \cdots + c_n^2)\sigma^2$ である．最良線形不偏推定量は，$c_1 = c_2 = \cdots = c_n$ であることが予想できる．不偏性より，$c_i = 1/n$，$\bar{c} = 1/n$ である．上記の不等式より，この推定量の分散 $n(1/n)^2\sigma^2 = (1/n)\sigma^2$ は任意の他の線形推定量の分散より大きくなることはない．

問 4.3 ポアソン分布での最尤推定： $Y_1, Y_2, \ldots, Y_n$ が互いに独立に，確率関数が次式で与えられる離散分布（ポアソン分布）に従うとき，λ の最尤推定量を求めよ．

$$f(y) = \frac{\lambda^y}{y!}\exp(-\lambda), \quad y = 0, 1, \cdots, \quad 0 < \lambda.$$

この最尤推定量が最良線形不偏推定量であることを証明せよ．

解答例： 尤度 $L(\lambda)$ は次の通り．

$$L(\lambda) = \frac{\lambda^{y_1+y_2+\cdots+y_n}}{y_1!y_2!\cdots y_n!}\exp(-n\lambda)$$

対数尤度を λ で微分して 0 とおいた解を $\hat{\lambda}$ とおくと次式が得られる．

$$\sum_{i=1}^{n} y_i/\hat{\lambda} - n = 0 \to \hat{\lambda} = \sum_{i=1}^{n} y_i/n = \bar{y}$$

したがって最尤推定量は $\bar{Y}$ である．

一般の線形推定量 $T = c_1Y_1 + c_2Y_2 + \cdots + c_nY_n$ が λ の不偏推定量であるためには　$E\{T\} = \lambda$ が必要である．その条件の下で分散最小のものを求めればよい．ポアソン分布では，平均も分散も λ であるから $c_1 + c_2 + \cdots + c_n = 1$ という条件の下で $V\{T\} = (c_1^2 + c_2^2 + \cdots + c_n^2)\lambda$ を最小にすればよい．問 4.2 で述べたように，この条件を満たすものは $c_1 = c_2 = \cdots = c_n = 1/n$ である．したがって最尤推定量は最良線形不偏推定量である．

問 4.4 最尤推定量の偏り： 4.3.3 節で求めた最尤推定量 $\hat{\mu}$ は不偏推定量であるが，(4.15) 式の $\hat{\sigma}^2$ は不偏推定量でないことを証明せよ．

解答例：

$$\begin{aligned}
E\{\hat{\mu}\} &= E\{\frac{1}{n}Y_1 + \frac{1}{n}Y_2 + \cdots + \frac{1}{n}Y_n\} \\
&= \frac{1}{n}E\{Y_1\} + \frac{1}{n}E\{Y_2\} + \cdots + \frac{1}{n}E\{Y_n\} \\
&= \frac{1}{n}\mu + \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu = \mu \\
E\{\hat{\sigma}^2\} &= E\{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\} = \frac{1}{n}E\{\sum_{i=1}^{n}[(Y_i - \mu) - (\bar{Y} - \mu)]^2\} \\
&= \frac{1}{n}E\{\sum_{i=1}^{n}(Y_i - \mu)^2\} - E\{(\bar{Y} - \mu)^2\} \\
&= \frac{n\sigma^2}{n} - E\{(\bar{Y} - \mu)^2\} = \frac{n-1}{n}\sigma^2 < \sigma^2
\end{aligned}$$

問 4.5 最小二乗推定量の最適性： 4.4 節の (4.24) 式で与えられる $\hat{\beta}_1$ が β_1 の線形不偏推定量であることを証明せよ．

（**本文の誤りの修正**　問題文中の「最良線形不偏推定量」は誤りで「線形不偏推定量」が正しい．）

解答例： $\sum_{i=1}^n(d_i-\bar{d})=0$ であることを利用すると，次のように線形性が示される．

$$\hat{\beta}_1 \;=\; \frac{\sum_{i=1}^n(d_i-\bar{d})(Y_i-\bar{Y})}{n\sum_{j=1}^n(d_j-\bar{d})^2}=\sum_{i=1}^n\frac{(d_i-\bar{d})}{n\sum_{j=1}^n(d_j-\bar{d})^2}Y_i$$

上の変形を利用して，不偏性も次のように示される．

$$\begin{aligned}
E\{\hat{\beta}_1\} &= \sum_{i=1}^n\frac{(d_i-\bar{d})}{n\sum_{j=1}^n(d_j-\bar{d})^2}E\{Y_i\}=\sum_{i=1}^n\frac{(d_i-\bar{d})}{n\sum_{j=1}^n(d_j-\bar{d})^2}n(\beta_0+\beta_1d_i)\\
&= \sum_{i=1}^n\frac{(d_i-\bar{d})}{n\sum_{j=1}^n(d_j-\bar{d})^2}n\beta_1d_i=\frac{\sum_{i=1}^n(d_i-\bar{d})d_i}{\sum_{j=1}^n(d_j-\bar{d})^2}\beta_1\\
&= \frac{\sum_{i=1}^n(d_i-\bar{d})^2}{\sum_{j=1}^n(d_j-\bar{d})^2}\beta_1=\beta_1
\end{aligned}$$

問 4.6 臨床試験： ある骨粗鬆症患者の集団から，ランダムに n 人を選んで臨床試験の被験者とする．治療開始から新規の椎体骨折が見出されるまでの期間（月数）を $Y_1,Y_2,\ldots,Y_n$ とし，これらは互いに独立で，確率分布が下の式（指数分布）で与えられる確率変数であるとする．

$$Pr\{Y_i>t\}=\exp\left(-\frac{t}{\theta}\right),\quad 0<t,\quad 0<\theta$$

Y_i の確率分布の密度関数を求め，期待値 $E\{Y_i\}$ を計算せよ．さらに $\boldsymbol{Y}=(Y_1,Y_2,\ldots,Y_n)$ にもとづく $E\{Y_i\}$ の最尤推定量を求め，その期待値と分散を計算せよ．

解答例： (9) 式で与えられているのは上側確率なので，これを t で微分して符号を逆にして，y_i を代入すると，Y_i の確率密度が y_i の関数という形式で次式になる．

$$\frac{1}{\theta}\exp\left(-y_i/\theta\right)$$

これに基づいて，積分変数を $u=y_i/\theta$ と変換して期待値を計算すると次の結果が得られる．

$$E\{Y_i\}=\int_0^\infty y_i\,\frac{1}{\theta}\exp\left(-y_i/\theta\right)dy_i=\theta\int_0^\infty u\exp\left(-u\right)\,du=\theta$$

$Y_1,Y_2,\ldots,Y_n$ が互いに独立なので，尤度 $L(\theta;\boldsymbol{Y})$ はそれぞれの確率密度の積として次式になる．

$$L(\theta,\boldsymbol{Y})=\prod_{i=1}^n\frac{1}{\theta}\exp\left(-\frac{Y_i}{\theta}\right)=\frac{1}{\theta^n}\exp\left(-\sum_{i=1}^n y_i/\theta\right)$$

この尤度の対数をとって θ で微分して 0 とおいた方程式を解くと次の最尤推定量が得られる．

$$\hat{\theta}=\sum_{i=1}^n Y_i/n$$

問 4.7 壺のモデル： 1 つの壺に 4 個の球が入っている．1 という数字の球は ν 個である．復元抽出を 10 回行って 1 という数字が 8 回出たときの ν の最尤推定値を求めよ．

解答例： ν は 0, 1, 2, 3, 4 のどれかであるが，0 と 4 はありえないから，残りの 3 個の値に対して，${}_{10}C_8(\nu/4)^8(1-\nu/4)^2$ を計算して，どれが最も大きい尤度を与えるか，調べればよい．答えは，$\nu = 3$ である．

問 4.8 ロジスティック関数の意味： 群の大きさが n の集団に対して，ある被験物質を用量 x で投与したときの死亡数を y とする．x が dx 増加したときの y の増加分 dy は y と生存率の積に比例することが考えられる．比例係数を β_1 とすると，次式が成り立つ．

$$\frac{dy}{dx} = \beta_1 \frac{n-y}{n} y$$

この微分方程式を解いて解がロジスティック関数になることを確かめよ．

（**ヒント：** 微分方程式を解くことが必要である．）

解答例： 次の計算によって，積分定数 β_0 を含めたロジスティック関数が得られる．

$$\begin{aligned}
\int \frac{1}{(n-y)y} dy &= \beta_1 \int dx \quad \rightarrow \quad \int (\frac{1}{n-y} + \frac{1}{y}) dy = \beta_1 x + \beta_0 \\
-\log(n-y) + \log(y) &= \beta_1 x + \beta_0 \quad \rightarrow \quad \frac{y}{n-y} = \exp(\beta_1 x + \beta_0) \\
y &= (n-y)\exp(\beta_1 x + \beta_0) \quad \rightarrow \quad \frac{y}{n} = \frac{\exp(\beta_1 x + \beta_0)}{1 + \exp(\beta_1 x + \beta_0)}
\end{aligned}$$

理解を確かめるための設問

問 5.1 ポアソン分布での最強力検定： a, b が与えられた正の数で，Y_1，Y_2 が互いに独立に，それぞれ強度 $a\lambda$，$b\lambda$ のポアソン分布に従うとき，ある与えられた値 λ_0 に関する次の仮説検定問題に対して，ネイマン・ピアソン流の検定方式を求めよ．

$$H_0 : \lambda = \lambda_0 \quad vs \quad H_1 : \lambda > \lambda_0$$

この検定方式が一様最強力検定であることを説明せよ．

解答例： λ_0 より大きいある値 λ_1 に対して尤度比 LR は次式になる．

$$\begin{aligned}
LR &= \left\{ \frac{\exp[-(a\lambda_0 + b\lambda_0)]}{y_1! y_2!} (a\lambda_0)^{y_1} (b\lambda_0)^{y_2} \right\} / \left\{ \frac{\exp[-(a\lambda_1 + b\lambda_1)]}{y_1! y_2!} (a\lambda_1)^{y_1} (b\lambda_1)^{y_2} \right\} \\
&= \exp[-(a+b)(\lambda_0 - \lambda_1)] (\frac{\lambda_0}{\lambda_1})^{y_1 + y_2}
\end{aligned}$$

これがある値以下であるときに仮説 $\lambda = \lambda_0$ を棄却するのがネイマン・ピアソン流の検定方式，すなわち，対立仮説 $\lambda = \lambda_1$ に対する最強力検定である．この LR は $y_1 + y_2$ の単調減少関数であるから，ある棄却限界値 C を用いて，$y_1 + y_2 > C$ ならば H_0 を棄却するという仮説検定方式と同じである．

C は λ_0 の値のみで定められ λ_1 に無関係であるから，$\lambda_1 > \lambda_0$ である任意の λ_1 に対して最強力検定である．すなわち，$H_1 : \lambda > \lambda_0$ に対する一様最強力検定である．

問 5.2 指数分布の尤度比検定： $Y_1, Y_2, \ldots, Y_n$ が互いに独立に，確率密度関数が次式で与えられる連続分布（指数分布）に従う確率変数であるとする．

$$f(y) = \frac{1}{\theta}\exp(-y/\theta), \quad 0 < y, \quad 0 < \theta$$

ある与えられた値 θ_0 についての次の仮説検定問題に対する尤度比検定を求めよ．

$$H_0 : \theta_0 \geq \theta \quad vs \quad H_1 : \theta_0 < \theta$$

解答例： 尤度は次式である．

$$L(\theta) = \frac{1}{\theta^n}\exp(-\frac{1}{\theta}\sum_{i=1}^{n} y_i)$$

これを最大にする θ の値は，$\sum_{i=1}^{n} y_i/n\ (= \bar{y})$ である．H_0 の下で尤度を最大にする θ の値は $\max(\theta_0, \bar{y})$ である．したがって，尤度比 LR は次式になる．

$$LR = \max\{1, (\frac{\bar{y}}{\theta_0})^n \exp[-n(\frac{\bar{y}}{\theta_0} - 1)]\}$$

尤度比は $\bar{y}$ の単調減少関数なので $\bar{Y} = \sum_{i=1}^{n} Y_i/n$ に対して，$Pr\{\bar{Y} > C|\theta = \theta_0\} = \alpha$ を満たすように定めた定数 C を用いて「$\bar{y} > C$ ならば H_0 を棄却する」としたものが有意水準 α の尤度比検定となる．

問 5.3 両側 t 検定： 5.5.4 節で与えられている $\hat{\sigma}_H^2$ と $\hat{\sigma}_E^2$ がそれぞれ最尤推定量であることを証明せよ．

解答例： (5.26) 式の対数をとることで，対数尤度 $\log L$ が次式となる．

$$\log L = -\frac{n_1 + n_2}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{(Y_{ij} - \mu_i)^2}{2\sigma^2}$$

これを μ_1, μ_2, σ^2 で偏微分すると次式が得られる．

$$\begin{aligned}
\frac{\log L}{\partial \mu_i} &= 2\sum_{j=1}^{n_i}\frac{(Y_{ij} - \mu_i)}{2\sigma^2}; \quad (i = 1, 2) \\
\frac{\log L}{\partial \sigma^2} &= -\frac{n_1 + n_2}{2\sigma^2} + \sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{(Y_{ij} - \mu_i)^2}{2\sigma^4}
\end{aligned}$$

これらの式を 0 とした方程式を解くと σ^2 の解として，$\hat{\sigma}_E^2$ が得られる．

$\mu_1 = \mu_2 = \mu$ の場合は偏微分が次式になる．

$$\frac{\log L}{\partial \mu} = 2\sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{(Y_{ij} - \mu)}{2\sigma^2}, \quad \frac{\log L}{\partial \sigma^2} = -\frac{n_1 + n_2}{2\sigma^2} + \sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{(Y_{ij} - \mu)^2}{2\sigma^4}$$

これらの式を 0 とした方程式を解くと σ^2 の解として，$\hat{\sigma}_H^2$ が得られる．

問 5.4 分散の式の誘導： 4.4 節の (4.24) 式で与えられている $\hat{\beta}_1$ について 5.6.3 節の (5.47) 式を証明せよ．

（**本文の誤りの修正**　(5.47）式の $\beta_0(1-\beta_0)$ は誤りで $\beta_0(1-\beta_0)/n$ が正しい．(5.48) 式の $\bar{P}(1-\bar{P})$ は誤りで $\bar{P}(1-\bar{P})/n$ が正しい.）

解答例： (4.24) 式の中央式の分子にある $\bar{Y}$ は消去しても差し支えないので，$\hat{\beta}_1$ の式は次のように変形できる．

$$\hat{\beta}_1 = \sum_{i=1}^{n} \frac{d_i - \bar{d}}{n\sum_{j=1}^{n}(d_j-\bar{d})^2} Y_i$$

これは $Y_1, Y_2, \cdots, Y_n$ の線形式で，帰無仮説の下で $V\{Y_i\} = n\beta_0(1-\beta_0)$; $(i = 1, 2, \ldots, n)$ であるから 3.6 節の公式 6 を適用することで次の結果が得られる．

$$\begin{aligned} V\{\hat{\beta}_1\} &= \sum_{i=1}^{n}(\frac{d_i - \bar{d}}{n\sum_{j=1}^{n}(d_j-\bar{d})^2})^2 n\beta_0(1-\beta_0) \\ &= \frac{1}{\sum_{j=1}^{n}(d_j-\bar{d})^2}\beta_0(1-\beta_0)/n \end{aligned}$$

問 5.5 フィーラー法による信頼区間： $Y_1, Y_2, \ldots, Y_n$ が互いに独立に $E\{Y_i\} = \beta_0 + \beta_1 x_i$，$V\{Y_i\} = \sigma^2$ の正規分布に従う確率変数であるとする．$x_1, x_2, \ldots, x_n$ は値が与えられている共変量であり，β_0 と β_1 は未知母数である．β_0 と β_1 の最小二乗推定量を簡単のために b_0, b_1 とすると，任意の値 θ に対して，$T = b_0 + b_1\theta$ は

$$E\{T\} = \beta_0 + \beta_1\theta$$

$$V\{T\} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \left(\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i\theta + n\theta^2 \right)\sigma^2$$

の正規分布に従う．

簡単のために，$V\{T\} = A\sigma^2$ と書く．$\beta_0 + \beta_1\theta = 0$ で θ を定義し，ある与えられた定数 θ_0 について，次の仮説検定問題を考える．

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

この問題に対する有意水準 5%の検定方式として次の検定が考えられる．

$$\frac{(b_0 + b_1\theta_0)^2}{A\ V} > 1.96^2 \quad \text{のとき}, H_0\text{を棄却する}.$$

ここで V は $\hat{\sigma}^2 = \sum_{i=1}^{n}(Y_i - b_0 - b_1 x_i)^2/(n-2)$ である．

この検定に基づいて，信頼水準 95% の θ の信頼区間を求めよ．

解答例： A は θ の関数なので　$A(\theta) = A_1\theta^2 - 2A_2\theta + A_3$ と書き，$\bar{x} = \sum_{i=1}^n x_i$ とすると，A_1, A_2, A_3 が次式となる．

$$A_1 = \frac{\sum_{i=1}^n x_i^2/n}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad A_2 = \frac{\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad A_3 = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

検定統計量に未知母数を入れてはいけないので，分母の A を $A(\theta_0)$ にすると，この仮説検定で有意にならない θ_0 は次式を満たすものとなる．

$$(b_1\theta_0 + b_0)^2 < 1.96^2(A_1\theta_0^2 - 2A_2\theta_0 + A_3)V$$

すなわち，次の 2 次不等式の解が区間になる場合に信頼水準 95% の信頼区間が得られる．

$$(b_1^2 - 3.84A_1V)\theta_0^2 + 2(b_0b_1 + 3.84A_2V)\theta_0 + (b_0^2 - 3.84A_3V) < 0$$

$b_1^2 - 3.84A_1V < 0$ のとき，この不等式の解は区間にならないので，信頼区間が作れない．そうでなければ θ について次の復号値を両端点に持つ信頼区間が得られる．

$$-\frac{b_0b_1 + 3.84A_2V}{b_1^2 - 3.84A_1V} \pm \frac{\sqrt{(b_0b_1 + 3.84A_2V)^2 - (b_1^2 - 3.84A_1V)(b_0^2 - 3.84A_3V)}}{b_1^2 - 3.84A_1V}$$

問 5.6 対立仮説の違い： $Y_1, Y_2, \ldots, Y_n$ が互いに独立に $N(\mu, 1)$ に従うとき，対立仮説によって検定方式が変わることを，次の 2 つの検定問題のそれぞれについての一様最強力検定を例にして，説明せよ．

$$H_0 : \mu = 0 \quad vs \quad H_1 : \mu > 0$$

$$H_0 : \mu = 0 \quad vs \quad H_1 : \mu < 0$$

（**問題文の誤りの訂正**　「尤度比検定」は誤りで「一様最強力検定」が正しい．）

解答例： 検定問題 $H_0 : \mu = 0 \quad vs \quad H_1 : \mu > 0$ に対しては 5.5.4 節で求めた検定で $\mu_0 = 0, \sigma_0 = 1$ としたものが一様最強力検定になる．すなわち,「$\bar{Y} > u(\alpha)$ ならば H_0 を棄却する」ものである．

もう 1 つの検定問題 $H_0 : \mu = 0 \quad vs \quad H_1 : \mu < 0$ に対しては，同様に「$\bar{Y} < -u(\alpha)$ ならば H_0 を棄却する」ものが一様最強力検定になる．同じ帰無仮説の検定でも，対立仮説が異なると最良の検定方式が異なることが分かる．

理解を確かめるための設問

問 6.1 外れ値のモデル： 6.2.3 節で，対立仮説の下での最尤推定量が (6.18) 式で与えられることを証明せよ．

解答例： $\mu_1, \mu_2, \cdots, \mu_n$ がすべて異なるときの対数尤度関数 $\log L$ は次式で与えられる．

$$\log L = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n}\frac{(Y_i - \mu_i)^2}{2\sigma^2}$$

σ^2 を固定した条件の下でこれを最大にするには，$\sum_{i=1}^{n}(Y_i - \mu_i)^2$ を最小にすればよい．

順序統計量，すなわち $Y_1, Y_2, \cdots, Y_n$ を小さい値から大きい値に並べたものを $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$ とおくと，$\hat{\mu} < \hat{\mu}_{(o)}$ という条件があれば，任意の $j \neq n$ に対して次式が成り立つ．

$$(Y_{(n)} - \hat{\mu})^2 + (Y_{(j)} - \hat{\mu}_{(o)})^2 \geq (Y_{(j)} - \hat{\mu})^2 + (Y_{(n)} - \hat{\mu}_{(o)})^2$$

この不等号関係は，j を固定して，$i \neq j, n$ についての二乗和を加えても変わらない．したがって，$\sum_{i=1}^{n}(Y_i - \mu_i)^2$ の最小値は，$\hat{\mu}_{(o)} = Y_{(n)}$ という条件の下で，$\sum_{i\neq n}(Y_{(i)} - \mu)^2$ を最小にするものである．すなわち $\hat{\mu} = \sum_{i=1}^{n-1} Y_{(i)}/(n-1)$ である．

$\hat{\mu}_{(o)}, \hat{\mu}$ をこのように求めた上で，$\log L$ 最大にする σ^2 を求めると，(6.18) 式が得られる．

問 6.2 出現率の最良線形不偏推定量： 6.3.1 節の (6.19) 式で与えられる統計量 T が π の最良線形不偏推定量であることを証明せよ．

解答例： 任意の線形推定量 $t(\boldsymbol{Y}) = c_1Y_1 + c_2Y_2 + \cdots + c_aY_a$ の期待値は次式である．

$$E\{t(\boldsymbol{Y})\} = c_1n_1\pi + c_2n_2\pi + \cdots + c_an_a\pi = (\sum_{i=1}^{a} n_ic_i)\pi$$

これが π に等しいための必要十分条件は $\sum_{i=1}^{a} n_ic_i = 1$ である．$t(\boldsymbol{Y})$ の分散は次式である．

$$V\{t(\boldsymbol{Y})\} = c_1^2n_1\pi(1-\pi) + c_2^2n_2\pi(1-\pi) + \cdots + c_a^2n_a\pi(1-\pi) = (\sum_{i=1}^{a} n_ic_i^2)\pi(1-\pi)$$

$n = n_1 + n_2 + \cdots + n_a$ とおいて $\sum_{i=1}^{a} n_i(c_i - 1/n)^2 \geq 0$ という式を展開し，不偏性の条件を適用すると次の不等式が得られる．

$$\sum_{i=1}^{a} n_ic_i^2 - 1/n \geq 0 \quad \rightarrow \quad \sum_{i=1}^{a} n_ic_i^2 \geq 1/n$$

等号が成り立つのは，$c_1 = c_2 = \cdots = c_a = 1/n$ のときである．ゆえに，分散 $(\sum_{i=1}^{a} n_ic_i^2)\pi(1-\pi)$ を最小にする線形不偏推定量は，$T = (\sum_{i=1}^{a} Y_i)/n$ である．

問 6.3 分散の最尤推定： 6.5.2 節において，分散が全て異なる場合と，全て等しい場合の分散の最尤推定量を求めよ．

（**問題文の訂正**　「6.5.2 節」は誤りで,「6.5.1 節」が正しい.）

解答例： 分散がすべて異なるときの対数尤度関数 $\log L$ は次式である.

$$\log L = \sum_{i=1}^{a}\left(-\frac{n_i}{2}\log(2\pi\sigma_i^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{n_i}(Y_{ij}-\mu_i)^2\right)$$

これを $\mu_1, \mu_2, \cdots, \mu_a, \sigma_1^2, \sigma_2^2, \cdots, \sigma_a^2$ に関して微分して 0 とおき，解を求めればよい．結果として $\sigma_1^2, \sigma_2^2, \cdots, \sigma_a^2$ の最尤推定量が次式となる.

$$\hat{\sigma}_i^2 = \sum_{j=1}^{n_i}(Y_{ij}-\hat{\mu}_i)^2/n_i; \quad (\hat{\mu}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i; \quad i = 1, 2, \ldots, a)$$

同様にして，$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2 = \sigma^2$ とおいたときの σ^2 の最尤推定量が次式となる．$\hat{\mu_i}$ は上式と同じである.

$$\hat{\sigma}^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(Y_{ij}-\hat{\mu}_i)^2/(\sum_{i=1}^{a} n_i)$$

問 6.4 ダネット検定とチューキー検定： ダネット検定の棄却限界値はチューキー検定の棄却限界値より絶対値が小さいことを説明せよ.

（**ヒント：** 片側検定と両側検定を別々に評価すること.）

解答例： 6.5.3 節で例示されている 3 群の枠組みの下で (6.68) 式の記号を使って説明する.

有意水準 α の片側検定の棄却限界値を，それぞれ C_D, C_T とすると，これらは次の式を満たさなければならない.

$$Pr\{\max(T_{21}, T_{31}) > C_D|H_0\} = \alpha, \quad Pr\{\max(T_{21}, T_{31}, T_{32}) > C_T|H_0\} = \alpha$$

$\max(T_{21}, T_{31}, T_{32}) \geq \max(T_{21}, T_{31})$ が成り立つので $C_T \geq C_D$ が成り立つ.

両側検定では，検定統計量が上記統計量の絶対値を取ったものになるが，その場合でも，検定統計量の不等号関係は変わらず，同じ不等号関係が成り立つ.

問 6.5 超幾何分布： 6.7 節の (6.81) 式において，$F_{11} + F_{12} = f_{1+}$ と $F_{11} + F_{21} = f_{+1}$ を与えた F_{11} の条件付き分布を求めよ.

解答例： 条件付き確率関数 $g(f_{11}|f_{11} + f_{12} = f_{1+}, f_{11} + f_{21} = f_{+1}, f_{11} + f_{12} + f_{21} + f_{22} = n)$ は，確率をすべて加えたときに 1 となるための定数 C を用いて次式で与えられる.

$$C\frac{n!}{f_{11}!(f_{1+}-f_{11})!(f_{+1}-f_{11})!(n-f_{1+}-f_{+1}+f_{11})!}(\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}})^{f_{11}}$$

一般の場合にこの式を簡単にすることは難しいが，$\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = 1$ が成り立っている場合には，定数 C を具体的に与えることができて，次式が得られる.

$$\begin{aligned} & g(f_{11}|f_{11} + f_{12} = f_{1+}, f_{11} + f_{21} = f_{+1}, f_{11} + f_{12} + f_{21} + f_{22} = n) \\ = \quad & \frac{f_{1+}!f_{2+}!f_{+1}!f_{+2}!}{n!f_{11}!(f_{1+}-f_{11})!(f_{+1}-f_{11})!(n-f_{1+}-f_{+1}+f_{11})!} \end{aligned}$$

理解を確かめるための設問

問 7.1 閉手順： $Y_1, Y_2, \ldots, Y_a$ は互いに独立にそれぞれ，平均が $\mu_1, \mu_2, \ldots, \mu_a$ で分散が同じ値 1 の正規分布に従うとする．$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_a$ であることが確かであるとする．

このとき，次の手順で逐次，検定を行って結論を出したとき，$H_0 : \mu_1 = \mu_2$ を誤って棄却する確率が 5% 以下になることを説明せよ．

手順 1. $Y_a - Y_1 < 1.64 \times \sqrt{2}$ であれば，H_0 を受容して手順を終了する．

手順 2. i を $a-1$ とする．

手順 3. $Y_i - Y_1 < 1.64 \times \sqrt{2}$ であれば，H_0 を受容して手順を終了する．

手順 4. $i = 2$ であれば H_0 を棄却して手順を終了する．

手順 5. i を 1 だけ差し引いて，すなわち $i-1 \to i$ として，手順 3 に戻る．

解答例： H_0 を誤って棄却する確率 P_I は次式である．

$$P_I = Pr\{Y_a - Y_1 \geq 1.64\sqrt{2}, Y_{a-1} - Y_1 \geq 1.64\sqrt{2}, \cdots, Y_2 - Y_1 \geq 1.64\sqrt{2} | H_0\}$$

これについて以下のように不等号関係を調べることで，第 1 種の過誤確率が 0.05 以下であることが示される．

$$\begin{aligned} P_I &= Pr\{\min(Y_a - Y_1, Y_{a-1} - Y_1, \cdots, Y_2 - Y_1) \geq 1.64\sqrt{2} | H_0\} \\ &\leq Pr\{Y_2 - Y_1 \geq 1.64\sqrt{2} | H_0\} = 0.05 \end{aligned}$$

問 7.2 固有科学的有意性： 本書では図 7.1 のデータについて 1 つの見解を述べているが，この見解が無条件に正しいというわけではない．このデータに含まれる情報の利用については，消費者，医学者，生産者，行政関係者，それぞれの立場での，異なる見解があると考えられる．どのような立場ではどのような情報の利用の仕方が考えられるか，検討せよ．

解答例： この種のデータには，文献「吉村功著，健康・栄養食品の研究における統計的データ解析の注意点，健康・栄養食品研究，11 巻 pp. 9 – 16，2008」に指摘されている問題点があることが多いが，それについては問題ないと仮定する．

消費者は，自分自身に費用に見合う利益があるかを考慮すべきであろう．

医学者は，健康状態の状況に照らして考えられる対処の選択肢としての意義を考えるべきであろう．

生産者は，このデータに含まれている情報の一般化可能性について，正確な情報の開示を考えるべきであろう．

行政関係者は，この情報を利用するときの注意点が消費者に正しくかつ十分に伝えることができるように，条件を整えるべきであろう．

問 7.3 決定樹の利用： 多群のデータの解析についての決定樹では，最初に等分散性の検定を行って，等分散でないことが確かめられたとき，ノンパラ手法の採用を勧めるものが多い．このような決定樹の妥当性と不当性を吟味せよ．

解答例： ノンパラ手法は一般に前提が不成立のときでも頑健であることが多い．しかしながらノンパラ手法は，たとえばすべてのデータが同じ確率分布に従う確率変数の観測値であるというような，それなりの前提の上で妥当性が保証されているものに過ぎない．決して万能ではない．前提がなぜ，どういう側面で成り立っていないかを探索的に吟味して，適切な変数変換を適用してみるというような，感度解析の視点を取り入れるべきであろう．

問 7.4 分散安定化変換： 7.6.2 節の分散安定化変換の微分方程式を解くことで 7.6.3 節の「z 変換」と「逆正弦変換」を誘導せよ．

（**ヒント**：微分方程式の解法を知らないと解けない問題である．）

解答例： 分散安定化変換の条件の (7.8) 式は次式となる．

$$\frac{df(\rho)}{d\rho}(1-\rho^2)=C$$

これを未知関数 f の満たすべき微分方程式として解関数 f を求めると以下のようになる．

$$\begin{aligned}
\int df &= C\int \frac{1}{1-\rho^2}d\rho \\
f(\rho) &= C\int \frac{1}{2}(\frac{1}{1-\rho}+\frac{1}{1+\rho})d\rho = \frac{1}{2}C(\int \frac{1}{1-\rho}d\rho + \int \frac{1}{1+\rho}d\rho) \\
&= \frac{1}{2}C[-\log(1-\rho)+\log(1+\rho)] = \frac{1}{2}C\left(\log\frac{1+\rho}{1-\rho}\right)
\end{aligned}$$

$C=1$ とおき，ρ を R に置き換えると次の分散安定化変換が得られる．

$$f(R)=\frac{1}{2}\log\frac{1+R}{1-R}$$

問 7.5 離散分布の p 値： ある薬物 A を飼料にまぜて 2 年間飼育したラットの発癌率が π であるとする．このラットを 3 匹飼育して 2 年後に屠殺したときの担癌動物数を Y とする．Y は，二項分布 $B(3,\pi)$ に従うものとする．Y を検定統計量とし，これが大きい値のとき発癌性があると判断するものとする．

帰無仮説が $H_0:\pi=0.1$ であるときの p 値の分布を求めよ．$\pi=0.5$ が真のときの p 値の分布を示せ．

（**ヒント** 検定統計量 T の分布が離散分布で，その実現値が t_o のとき，p 値は $Pr\{T\geq t_o|H_0\}$ で定義するのがふつうである．）

解答例： Y の取り得る値は $\{0,\ 1,\ 2,\ 3\}$ で，その確率分布は $Pr\{Y=y\}={}_3C_y\pi^y(1-\pi)^{3-y}$ である．$H_0:\pi=0.1$ の下では $Pr\{Y\geq y\}$ が次の値になる．

$$Pr\{Y\geq 0\}=1.0000, Pr\{Y\geq 1\}=0.02710, Pr\{Y\geq 2\}=0.0280, Pr\{Y\geq 3\}=0.0000$$

すなわち，p 値の取り得る値は，小数点以下 4 桁で評価したとき，$\{1.000, 0.2710, 0.0280, 0.0010\}$ であり，それぞれの値を取る確率は，$\{0.7290, 0.2430, 0.0270, 0.0010\}$ である $\pi=0.5$ のときの確率変数としての P 値の確率分布は，$\pi=0.5$ のときに Y が 0, 1, 2, 3 の各値を取る確率になるから，次の通りになる．

$$Pr\{P=1.0000\}=0.1250, Pr\{P=0.2710\}=0.3750, Pr\{P=0.0280\}=0.3750, Pr\{P=0.0010\}=0.1250$$

p 値が 0.05 以下であれば H_0 を棄却するという仮説検定を採用したとき，対立仮説 $H_1=\pi=0.5$ における検出力は，0.8750 になるわけである．