演習問題解答

演習 1　ある競走馬がトレーニングをした後で，心拍数が 1 分間に 100 回以下になるまでの時間を 16 回測定したところ，平均で 490 秒で，不偏分散が 8100 であった．競走馬の平均的な回復時間 530 秒より短いと判断できるか，有意水準 5% で検定せよ．また，有意水準 1% の場合はどうなるか検討せよ．

解答　対象の競走馬の回復時間が $N(\mu, \sigma^2)$ に従っているものとすると，μ がその競走馬の平均回復時間である．したがって，他の競走馬の平均回復時間 $\mu_0 = 530$ に対して，帰無仮説 $H_0 : \mu = \mu_0 = 530$ と対立仮説 $H_1 : \mu < \mu_0 = 530$ の仮説検定を行うことになる．対象の競走馬の回復時間の測定値 $X_1, \ldots, X_n$ に対して，標本平均と標本不偏分散を $\bar{X}, U^2$ とすると，どちらの仮説が正しい時でも，$T := \sqrt{n}(\bar{X}-\mu)/\sqrt{U^2}$ は自由度 $n-1$ のティー分布に従うので，有意水準 α の棄却域 $R = \{T_0 \mid T_0 < -t_{n-1}(\alpha)\}$ となる．ここで，T_0 は $\mu = \mu_0$ にした時の T，つまり，$T_0 := \sqrt{n}(\bar{X} - \mu_0)/\sqrt{U^2}$ である．この問題では $n = 16$, $T_0 = -16/9 = -1.77778 < -t_{15}(0.05) = -1.753$ なので，有意水準が 5% では H_0 が棄却され，530 秒より有意に小さいと言える．また，$T_0 = -1.77778 > -t_{15}(0.01) = 12.602$ だから，有意水準が 1% では H_0 が棄却されず，530 秒より有意に小さいとは言えない．

演習 2　50 歳代サラリーマンに老後の不安を調査したところ，192 人中 144 人が不安があると答えた．日本全体の 50 歳代サラリーマンの老後に不安があると思っている人の割合について，99% 両側信頼区間を求めよ．また，95% 上側信頼区間を求めよ．

解答　50 歳代サラリーマン全体における，老後の不安を感じる人の割合を p とすると，無作為に選んだ n 人中不安をあると答える人の人数 X は二項分布 $B(n,p)$ に従う．ただし，n が大きい (およそ 100 以上) の時は，$\bar{X} = X/n$ とすると，$Z := \sqrt{n}(\bar{X}-p)/\sqrt{\bar{X}(1-\bar{X})}$ は近似的に標準正規分布 $N(0,1)$ に従う (中心極限定理)．したがって，信頼度 $1-\alpha$ の両側信頼区間は

$$\left[\bar{X} - z(\alpha/2)\sqrt{\frac{\bar{X}(1-\bar{X})}{n}},\ \ \bar{X} + z(\alpha/2)\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]$$

で与えれる．この問題では，$\bar{X} = 144/192 = 3/4, n = 192$ であり，$z(0.005) = 2.575$ なので，99% 両側信頼区間は

$$\left[3/4 - 2.575 \times \sqrt{\frac{(3/4)(1-3/4)}{192}},\ \ 3/4 + 2.575 \times \sqrt{\frac{(3/4)(1-3/4)}{192}}\right] = [0.670,\ \ 0.830]$$

である．一方，信頼度 $1-\alpha$ の上側信頼区間は

$$\left[\bar{X} - z(\alpha)\sqrt{\frac{\bar{X}(1-\bar{X})}{n}},\ \ 1\right]$$

で与えれるので，$z(0.05) = 1.645$ だから，95% 上側信頼区間は，

$$\left[3/4 - 1.645 \times \sqrt{\frac{(3/4)(1-3/4)}{192}},\quad 1\right] = [0.700,\quad 1]$$

となる．

演習 3　$X_1, \ldots, X_n$ が平均 μ の指数分布に従う無作為標本であるとする．そのとき，$2n\bar{X}/\mu \sim \chi^2_{2n}$ が知られている．このことを利用して，μ の信頼度 $1-\alpha$ の両側信頼区間を作れ．また，

$$\begin{cases} H_0 & : \ \mu = \mu_0 \\ H_1 & : \ \mu > \mu_0 \end{cases}$$

に対する有意水準 α の棄却域 R を作れ．

解答

《信頼区間》

$$P\left\{\chi^2_{2n}(1-\alpha/2) \leqq 2n\bar{X}/\mu \leqq \chi^2_{2n}(\alpha/2)\right\} = 1-\alpha$$

なので，信頼度 $1-\alpha$ の両側信頼区間は

$$\left[\frac{2n\bar{X}}{\chi^2_{2n}(\alpha/2)},\quad \frac{2n\bar{X}}{\chi^2_{2n}(1-\alpha/2)}\right]$$

《仮説検定・棄却域》まず棄却域 R の形を決める．$H_1 \ : \ \mu > \mu_0 \iff \mu - \mu_0 > 0 \iff \mu/\mu_0 > 1$, $\mu \approx \bar{X}$ であり，$H_0 \ : \mu = \mu_0$ が正しい時，$V_0 := 2n\bar{X}/\mu_0 \sim \chi^2_{2n}$ であることを考慮にいれると，H_1 と判定するための条件として，$\bar{X}/\mu_0 > A$ という形を採用すればよいと考えられる．つまり棄却域の形は，$R = \{\bar{X} \mid \bar{X}/\mu_0 > A\}$ とすればよいことがわかる．この判定条件を用いる時，H_0 が正しいのに H_1 と判定する誤りの確率は

$$P\{\bar{X}/\mu_0 > A\} = P\{2n\bar{X}/\mu_0 > 2nA\} = P\{V_0 > 2nA\}$$

で与えられる．したがって，その確率が α であるとき，

$$P\{V_0 > 2nA\} = \alpha$$

なので，$2nA = \chi^2_{2n}(\alpha)$ とすればよいことがわかる．以上より，棄却域 $R = \{\bar{X} \mid V_0 = 2n\bar{X}/\mu_0 > \chi^2_{2n}(\alpha)\}$ とすればよいことが分かった．

演習 4　《例 2.1》 で考えた血圧降下剤に関する測定値 Z_i が独立で同一の $N(\mu, \sigma^2)$ 従うとして，以下の問に答えよ．.

(1)　平均降下量 μ の $100(1-\alpha)\%$ 信頼区間を求めよ．

(2) $\mu < -10$ かどうかは判定したい．帰無仮説 H_0 と対立仮説 H_1 を設定して，有意水準 α の棄却域 R をスチューデント化変量 T の領域として表せ．ただし，自由度 ν のティー分布の上側 $100\alpha\%$ 点を $t_\nu(\alpha)$ と表すことにする．

(3) 10 人に対する測定値から $\bar{Z} = -1\ 4$, $U_Z^2 = 3$ をえた．これから μ の 95% 信頼区間を構成せよ．また，上の仮説検定を有意水準 5% で実行せよ．

解答

(1) $[\bar{Z} - t_{n-1}(\alpha/2)\frac{U_Z}{\sqrt{n}}, \bar{Z} + t_{n-1}(\alpha/2)\frac{U_Z}{\sqrt{n}}]$.

(2) $R = \{T \,|\, T < -t_{n-1}(\alpha)\}$. $T = \sqrt{n}(\bar{Z} + 10)/U_Z$.

(3) $\bar{Z} \pm t_{10-1}(0.025) \times U_Z/\sqrt{n} = -14 \pm 2.262 \times \sqrt{3}/\sqrt{10} = -15.2389 \sim -12.7611$. また，$T = \sqrt{10}(-14 + 10)/\sqrt{3} = -7.30297 < -t_{10-1}(0.05) = -1.833$ なので，対象の薬剤の効果は有意である．

演習 5 《例 2.2》 で考えた 2 つの中学の平均遠投距離 μ_A, μ_B に関する測定値 $X_i, i = 1, \ldots, m$ と $Y_j, j = 1, \ldots, n$ が $N(\mu_A, \sigma_A^2)$, $N(\mu_B, \sigma_B^2)$ 従うとして，以下の問に答えよ..

(1) $\sigma_A^2 = \sigma_B^2$ のとき，帰無仮説 $H_0 : \mu_A = \mu_B$，対立仮説 $H_1 : \mu_A \neq \mu_B$ に対する有意水準 α の棄却域 R をスチューデント化変量 T の領域として表せ．

(2) $\sigma_A^2 \neq \sigma_B^2$ のとき，帰無仮説 $H_0 : \mu_A = \mu_B$，対立仮説 $H_1 : \mu_A \neq \mu_B$ に対する有意水準 α の棄却域 R をスチューデント化変量 T の領域として表せ．

(3) 帰無仮説 $H_0 : \sigma_A^2 = \sigma_B^2$，対立仮説 $H_1 : \sigma_A^2 \neq \sigma_B^2$ に対する有意水準 α の棄却域 R を $R := U_x^2/U_y^2$ の領域として表せ．ただし，自由度 (ν_1, ν_2) のエフ分布の上側 $100\alpha\%$ 点を $f_{\nu_1,\nu_2}(\alpha)$ と表すことにする．

解答

(1) $R = \{T \,|\, |T| > t_{m+n-2}(\alpha/2)\}$, $T = \frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)\hat{\sigma}^2}}$, $\hat{\sigma}^2 = \frac{(m-1)U_x^2+(n-1)U_y^2}{m+n-2}$.

(2) $R = \{T \,|\, |T| > t_{\hat{c}}(\alpha/2)\}$, $T = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{1}{m}U_x^2+\frac{1}{n}U_y^2}}$, $\hat{c} = \frac{1}{\frac{\hat{d}^2}{m-1}+\frac{(1-\hat{d})^2}{n-1}}$, $\hat{d} = \frac{\frac{U_x^2}{m}}{\frac{U_x^2}{m}+\frac{U_y^2}{n}}$.

(3) $R = \{U_x^2/U_y^2 \,|\, U_x^2/U_y^2 < f_{m-1,n-1}(1-\alpha/2), f_{m-1,n-1}(\alpha/2) < U_x^2/U_y^2\}$.

演習 6 ある年のセンター試験の英語とその追試験の平均点が公表されていない時，その違いを調べるため無作為に選んだ受験生から聞き取り調査をした．その結果，センター試験の受験生 10 人の平均点は 125 点で標準偏差が 24 点，追試験の受験生 10 人の平均点が 100 点で標準偏差は 45 点であった．センター試験とその追試験の得点分布がそれぞれ，$N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ であるとして，以下の問いに答えよ．

(1) 設定すべき帰無仮説 H_0 と対立仮説 H_1 を μ_1, μ_2 を用いて表せ．

(2) $\sigma_1^2 = \sigma_2^2$ として，有意水準 5% で仮説検定しなさい．

(3) σ_1^2 と σ_2^2 が等しいかどうか，有意水準 10% で仮説検定しなさい．

解答 この問題は 2 つの正規母集団の差に関するもので，対応のない場合の平均の差の問題である．

まず，本試験の平均点 μ_1 と追試験の平均点 μ_2 の違いを検討するので，$H_0 : \mu_1 = \mu_2$，$H_1 : \mu_1 \neq \mu_2$ であることがわかる．

無作為に選んだ本試験の受験生 m 人の点数を $X_1, \ldots, X_m$，無作為に選んだ追試験の受験生 n 人の点数を $Y_1, \ldots, Y_n$ として，$U_x^2 = \frac{1}{m-1}\sum_{i=1}^m (X_i - \bar{X})^2$, $U_y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$，$\hat{\sigma}^2 = \frac{(m-1)U_x^2 + (n-1)U_y^2}{m+n-2}$ とすると，$\sigma_1^2 = \sigma_2^2 =: \sigma^2$ の時，

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\hat{\sigma}^2}} \sim t_{m+n-2}$$

であることを利用する．棄却域 R は，$H_1 : \mu_1 \neq \mu_2 \iff |\mu_1 - \mu_2| \neq 0$ と $\mu_1 - \mu_2 \approx \bar{X} - \bar{Y}$ を考慮にいれると, $R = \{(\bar{X}, \bar{Y}) \mid |\bar{X} - \bar{Y}| > A\}$ であることがわかる．$H_0 : \mu_1 = \mu_2$ が正しい時，

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\hat{\sigma}^2}} =: T_0 \sim t_{m+n-2}$$

なので，H_0 が正しいにもかかわらず H_1 が正しいと判定する確率 α は

$$\begin{aligned}
\alpha =& P\{|\bar{X} - \bar{Y}| > A\} \\
=& P\left\{\frac{|\bar{X} - \bar{Y}|}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\hat{\sigma}^2}} > \frac{A}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\hat{\sigma}^2}}\right\} \\
=& P\{|T_0| > B\}
\end{aligned}$$

と表せることがわかる．ここで，

$$B = \frac{A}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\hat{\sigma}^2}}.$$

したがって，$B = t_{m+n-2}(\alpha/2)$ となり，結局

$$R = \left\{ (\bar{X}, \bar{Y}) \;\middle|\; T_0 = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\hat{\sigma}^2}} > t_{m+n-2}(\alpha/2) \right\}$$

となる．

いま，$m = n = 10$, $\bar{X} = 125$, $\bar{Y} = 100$, $U_x^2 = 10 \times 24^2/9 = 640$, $U_y^2 = 10 \times 45^2/9 = 2250$ なので，$\hat{\sigma}^2 = 1445$ であり，

$$T_0 = \frac{125 - 100}{\sqrt{\left(\frac{1}{10} + \frac{1}{10}\right) \times 1445}} = \frac{25}{17} = 1.47059$$

である．$t_{18}(0.025) = 2.101$ だから，$|T_0| < t_{18}(0.025)$ なので，H_0 は棄却されない．つまり，本試験と追試験の平均点に有意な差はない．

最後に，$H_0 \ : \ \sigma_1^2 = \sigma_2^2$ と $H_1 \ : \ \sigma_1^2 \neq \sigma_2^2$ について考える．ただし，$W = (U_x^2/\sigma_1^2)/(U_y^2/\sigma_2^2) \sim F_{m-1,n-1}$ を利用することにする．棄却域 R の形は，$H_1 \ : \ \sigma_1^2 \neq \sigma_2^2 \iff \sigma_1^2/\sigma_2^2 \neq 1$ と $\sigma_1^2 \approx U_x^2$, $\sigma_2^2 \approx U_y^2$ を考慮に入れると，$R = \{(U_x^2, U_y^2) \mid U_x^2/U_y^2 < A_1,\ A_2 < U_x^2/U_y^2\}$ であることがわかる．今，$H_0 \ : \ \sigma_1^2 = \sigma_2^2$ が正しい時，

$$W = \frac{U_x^2}{U_y^2} =: W_0 \sim F_{m-1,n-1}$$

なので，H_0 が正しい時に H_1 が正しいと判定する確率 α は

$$\begin{aligned} \alpha =& P\{U_x^2/U_y^2 < A_1,\ A_2 < U_x^2/U_y^2\} \\ =& P\{W_0 < A_1,\ A_2 < W_0\} \end{aligned}$$

と表せて，$A_1 = f_{m-1,n-1}(1 - \alpha/2)$, $A_2 = f_{m-1,n-1}(\alpha/2)$ と選べばよいことがわかる．以上より，

$$R = \{(U_x^2, U_y^2) \mid W_0 = U_x^2/U_y^2 < f_{m-1,n-1}(1 - \alpha/2),\ f_{m-1,n-1}(\alpha/2 < U_x^2/U_y^2 = W_0\}$$

となる．さて，この問題では，$m = n = 10$, $W_0 = 640/2250 = 0.284444$ であり，$f_{9,9}(0.05) = 3.18$ である．また，一般に $f_{\nu_1,\nu_2}(1-\alpha) f_{\nu_2,\nu_1}(\alpha) = 1$ が成り立つので，$f_{9,9}(0.95) = 1/f_{9,9}(0.05) = 0.314465$ がわかる．したがって，$W_0 < f_{9,9}(0.95)$ が成り立ち，H_0 が棄却される．つまり，$\sigma_1^2 = \sigma_2^2$ とは言えない．この結果より，(2) で $\sigma_1^2 = \sigma_2^2$ と仮定したことは誤りであることがわかった．

演習 7　あるタレントの好感度に関して，無作為に選んだ n 人から聞き取り調査をした結果，x_1 人が好き，x_2 人が嫌い，x_3 人がどちらでもないと答えるものとする．

(1) 好きな人と嫌いな人の比率の差について，95% 両側信頼区間を作れ．

(2) 嫌いな人が好きな人より多いかどうか検定したい．帰無仮説 H_0 と対立仮説 H_1 を，好きな人の比率 p_1 と嫌いな人の比率 p_2 用いて表せ．さらに，有意水準 α の棄却域 R を作れ．

解答　この問題は，同一集団に対する比率の差に関するものである．したがって，集団全体において好きな人の割合を p_1, 嫌いな人の割合を p_2 として，n に調査した結果の割合をそれぞれ $\bar{x}_1 = x_1/n$, $\bar{x}_2 = x_2/n$ とおくと，n が大きい時，

$$Z = \sqrt{n}\frac{\bar{x}_1 - \bar{x}_2 - (p_1 - p_2)}{\sqrt{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}}$$

は近似的に標準正規分布 $N(0,1)$ に従う．したがって，

$$P(|Z| \leqq 1.96) = 0.95$$

が成り立つので，95% 両側信頼区間は $|Z| \leqq 1.96$ を $p_1 - p_2$ について解いて，

$$\bar{x}_1 - \bar{x}_2 - 1.96 \times \sqrt{\frac{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}{n}} \leqq p_1 - p_2 \leqq \bar{x}_1 - \bar{x}_2 + 1.96 \times \sqrt{\frac{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}{n}}$$

つまり，

$$\left[\bar{x}_1 - \bar{x}_2 - 1.96 \times \sqrt{\frac{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}{n}},\ \ \bar{x}_1 - \bar{x}_2 + 1.96 \times \sqrt{\frac{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}{n}}\right]$$

である．

また,『嫌いな人が好きな人より多い』を数式で表すと $p_1 < p_2$ であるが，これを検定するために，$H_0 : p_1 = p_2$, $H_1 : p_1 < p_2$ とする．この仮説検定のための棄却域 R の形を考える．まず，$H_1 : p_1 < p_2 \iff p_1 - p_2 < 0$ であることと $p_1 - p_2 \approx \bar{x}_1 - \bar{x}_2$ を考慮に入れると，$R = \{(\bar{x}_1, \bar{x}_2) \mid \bar{x}_1 - \bar{x}_2 < -A\}$ という形であることがわかる．さらに，$H_0 : p_1 = p_2$ の下で，

$$Z = \sqrt{n}\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}} =: Z_0$$

は近似的に標準正規分布 $N(0,1)$ に従うことが分かっているので，H_0 が正しいのに H_1 と判定する確率 α は，

$$\begin{aligned}\alpha =& P\{\bar{x}_1 - \bar{x}_2 < -A\} \\ =& P\left\{\sqrt{n}\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}} < -\sqrt{n}\frac{A}{\sqrt{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}}\right\} \\ =& P\{Z_0 < -B\}\end{aligned}$$

と表せる．ここで，

$$B = \sqrt{n}\frac{A}{\sqrt{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}}$$

とおいた．したがって，$B = z(\alpha)$ であることが導かれる．以上より，

$$R = \left\{ (\bar{x}_1, \bar{x}_2) \;\middle|\; \sqrt{n}\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\bar{x}_1 + \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)^2}} \right\}$$

演習 8

(1) ある実験の結果は k 種類であり，それぞれの結果が起こる確率は p_1, ..., p_k であるとする．このとき，帰無仮説 $H_0 : p_1 = a_1$, ..., $p_k = a_k$，対立仮説 $H_1 : (H_0$でない$)$ に対する仮説検定を行いたい．実験を n 回繰り返した時，k 種類の結果の出た回数を X_1, ..., X_k とする．このとき，有意水準 α の棄却域を構成せよ．ただし，自由度 ν のカイ 2 乗分布の上側 $100\alpha\%$ 点 $\chi^2_\nu(\alpha)$ を用いること．

解答　$n = x_1 + \cdots + x_k$ とすると，$P(X_1 = x_1, \ldots, X_k = x_k) = n!\dfrac{p_1^{x_1}}{x_1!}\cdots\dfrac{p_k^{x_k}}{x_k!}$ であり，n が大きい時，

$$V = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}$$

の分布は近似的に自由度 $k-1$ のカイ 2 乗分布に従う．さて，H_1 は $p_i \neq a_i \iff |p_i - a_i| \neq 0$ となる i が存在することであるが，$npi \approx X_i$ であること考慮すると，$R = \{(X_1, \ldots, X_k) \mid |X_1 - na_1| > A_1, \ldots, |X_k - na_k| > A_k\}$ とすればよいことがわかる．ところで，H_0 が正しい時，

$$V = \sum_{i=1}^{k} \frac{(X_i - na_i)^2}{na_i} =: V_0 \sim \chi^2_{k-1}$$

なので，この形に注目して，

$$R = \left\{ (X_1, \ldots, X_k) \;\middle|\; V_0 = \sum_{i=1}^{k} \frac{(X_i - na_i)^2}{na_i} > \sum_{i=1}^{k} \frac{A_i^2}{na_i} \right\}$$

とすることにする．簡単のため，

$$B = \sum_{i=1}^{k} \frac{A_i^2}{na_i}$$

とおく．H_0 が正しい時 H_0 を棄却する確率 α は，

$$\alpha = P\{V_0 > B\}$$

であり，$V_0 \sim \chi^2_{k-1}$ なので，$B = \chi^2_{k-1}(\alpha)$ とすればよいことがわかる．よって，

$$R = \left\{ (X_1, \ldots, X_k) \;\middle|\; V_0 = \sum_{i=1}^{k} \frac{(X_i - na_i)^2}{na_i} > \chi^2_{k-1}(\alpha) \right\}.$$

(2) ある神社で正月におみくじを引くと，大凶はほとんどでないといわれている．おみくじの種類は大吉，中吉，吉，大凶の 4 種類であるとき，★「大凶が大吉の 10 分の 1 で，中吉，吉の 5 分の 1」であるかどうか検定するものとする．おみくじを引いた人を 100 人無作為に選んで，何が出たかを調査したところ，大吉 40 人，中吉 30 人，吉 28 人，大凶 7 人 であった．この結果から，★の仮説が正しいかどうか検定せよ．有意水準は，5% と 1% で．

解答　大吉，中吉，吉，大凶が出る確率を p_1, p_2, p_3, p_4 とする．仮説★の下で，$p_4 = p_1/10 = p_2/5 = p_3/5$ なので，$p_1 + \cdots + p_4 = 1$ を考慮すると，$p_1 = 10/21$, $p_2 = 5/21$, $p_3 = 5/21$, $p_4 = 1/21$ となる．したがって，$a_1 = 10/21$, $a_2 = 5/21$, $a_3 = 5/21$, $a_4 = 1/21$ とする．いま，$n = 105$ なので，$V_0 = 104/25 = 4.16$. $\chi^2_3(0.05) = 7.815$, $\chi^2_3(0.01) = 11.345$ なので，どちらの有意水準でも H_0 は棄却されない．

演習 9　$\sqrt{2}$ の小数第 999 ケタまでに含まれる数字の度数を調べると，下の表のようになった．0 から 9 までの数字が同じ割合で現れるかどうか有意水準 5% で検定せよ．

数字	0	1	2	3	4	5	6	7	8	9
度数	108	99	108	82	100	104	90	104	113	92

解答　数字 i が現れる確率を p_i, $i = 0, 1, \ldots, 9$ とすると,『同じ割合で現れる』ことを帰無仮説 H_0 とすると，

$$H_0 \;\; p_0 = p_1 = \cdots = p_9 \iff p_0 = \frac{1}{10},\ p_1 = \frac{1}{10}, \ldots,\ p_9 = \frac{1}{10}$$

となる．演習 7 より，

$$\begin{aligned} V_0 =& \frac{(108-100)^2}{100} + \frac{(99-100)^2}{100} + \frac{(108-100)^2}{100} + \frac{(82-100)^2}{100} + \frac{(100-100)^2}{100} \\ &+ \frac{(104-100)^2}{100} + \frac{(90-100)^2}{100} + \frac{(104-100)^2}{100} + \frac{(113-100)^2}{100} + \frac{(92-100)^2}{100} \\ =& \frac{818}{100} = 8.18 < \chi^2_9(0.05) = 16.919 \end{aligned}$$

したがって，同じ割合で現れないとは言えない．

演習 10 アメリカ少数民族から無作為に選ばれた 550 の血液型を調査した結果，下の表のようになった．民族の違いと血液型は独立であるかどうか，有意水準 1% で検定せよ．ただし，表中の括弧内の数字は，独立であると仮定したときの期待度数である．

	A	B	O	AB	計
スペイン系	105(103.1)	60(59.89)	90(83.45)	15(23.56)	270
オリエント系	50(60.32)	40(35.05)	42(48.84)	26(13.79)	158
他のアジア系	55(46.58)	22(27.06)	38(37.71)	7(10.65)	122
計	210	122	170	48	550

解答 二つの分類基準 A, B において，$A_1, \ldots, A_r, B_1, \ldots, B_s$ という項目に分類し，下のような表が得られたとする．

	B_1	B_2	$\cdots$	B_s	計
A_1	X_{11}	X_{12}	$\cdots$	X_{1s}	$X_{1\bullet}$
A_2	X_{21}	X_{22}	$\cdots$	X_{2s}	$X_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_r	X_{r1}	X_{r2}	$\cdots$	X_{rs}	$X_{r\bullet}$
計	$X_{\bullet 1}$	$X_{\bullet 2}$	$\cdots$	$X_{\bullet s}$	n

分類項目 A_i, B_j に含まれる母集団の要素の数の割合を p_{ij} とすると，$X_{11} = x_{11}, \ldots, X_{rs} = x_{rs}$ となる確率は

$$P(X_{11} = x_{11}, \ldots, X_{1s} = x_{1s}, \ldots, X_{r1} = x_{r1}, \ldots, X_{rs} = x_{rs}) = n! \prod_{i=1}^{r} \prod_{j=1}^{s} \frac{p_{ij}^{x_{ij}}}{x_{ij}!}$$

となる．ただし，

$$\sum_{i}^{r} \sum_{j}^{s} X_{ij} = n, \quad \sum_{i}^{r} \sum_{j}^{s} p_{ij} = 1$$

基準 A と B が独立であるという仮説を帰無仮説 H_0 とすると，

H_0 : $p_{ij} = p_{i\cdot} p_{\cdot j}$, $i = 1, \ldots, r$, $j = 1, \ldots, s$ となる．帰無仮説 H_0 が真のとき，$\hat{p}_{ij} = X_{i\cdot} X_{\cdot j}/n^2$ とすると，

$$V_0 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(X_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \overset{d}{\Rightarrow} \chi^2_{(r-1)(s-1)}, \quad (n \to \infty)$$

が知られているので，対立仮説 H_1 : (H_0でない) に対する有意水準 α の棄却域として，

$$R = \left\{ (X_{11}, \ldots, X_{rs}) \;\middle|\; V_0 > \chi^2_{(r-1)(s-1)}(\alpha) \right\}$$

とすればよい．

さて，この問題では，$r=3,\ s=4,\ n=12$ であり，$\hat{p}_{ij}$ は表中の括弧内に与えられているので，

$$V_0 = \frac{(105-103.1)^2}{103.1} + \cdots + \frac{(7-10.65)^2}{10.65} = 21.62 > \chi_6^2(0.01) = 16.812$$

となり，民族の違いと血液型は独立でないことがわかる．

演習 11　データ $(x_1, y_1), \ldots, (x_n, y_n)$ の回帰直線を $y = \hat{\alpha}x + \hat{\beta}$ として，それによる $x = x_i$ に対する y の予測値を $\hat{y}_i$，残差を $\hat{e}_i$ とする．以下の問いに答えよ．

(1) $\bar{y} = \frac{1}{n}\sum_{i=1}^n \hat{y}_i$ を示せ．

解答　$\hat{y}_i = \hat{\alpha}x_i + \hat{\beta},\ \hat{\beta} = \bar{y} - \hat{\alpha}\bar{x}$ だから，$\hat{y}_i = \hat{\alpha}(x_i - \bar{x}) + \bar{y}$. したがって，

$$\frac{1}{n}\sum_{i=1}^n \hat{y}_i = \hat{\alpha}\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}) + \frac{1}{n}\sum_{i=1}^n \bar{y} = \bar{y}.$$

(2) $S_{ee} = n\left(s_y^2 - \dfrac{s_{xy}^2}{s_x^2}\right)$ を示せ．

解答　$\hat{e}_i = y_i - (\hat{\alpha}x_i + \hat{\beta}) = y_i - \bar{y} - \hat{\alpha}(x_i - \bar{x})$ なので，$\hat{y}_i = \hat{a}(x_i - \bar{x}) + \bar{y}$ だから，

$$\begin{aligned} S_{ee} =& \sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (y_i - \bar{y})^2 - 2\hat{\alpha}\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) + \hat{\alpha}^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ =& n(s_y^2 - 2\hat{\alpha}s_{xy} + \hat{\alpha}^2 s_x^2) = n\left(s_y^2 - 2\frac{s_{xy}}{s_x^2}s_{xy} + \frac{s_{xy}^2}{s_x^4}s_x^2\right) = n\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right). \end{aligned}$$

演習 12　$s_{xy} \sim N\left(\alpha s_x^2, \dfrac{s_x^2}{n}\sigma^2\right)$ を示せ．

解答　$s_{xy} = \displaystyle\sum_{i=1}^n p_i y_i,\ p_i = \frac{x_i - \bar{x}}{n},\ y_i \sim N(\mu_i, \sigma^2),\ \mu_i = \alpha x_i + \beta$ だから，

$$s_{xy} \sim N\left(\sum_{i=1}^n p_i\mu_i, \sum_{i=1}^n p_i^2\sigma^2\right)$$

である．ここで，

$$\begin{aligned}\sum_{i=1}^{n} p_i\mu_i =& \sum_{i=1}^{n} \frac{x_i-\bar{x}}{n}(\alpha x_i+\beta) = \frac{\alpha}{n}\sum_{i=1}^{n}(x_i-\bar{x})x_i + \frac{\beta}{n}\sum_{i=1}^{n}(x_i-\bar{x}) \\ =&\frac{\alpha}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x}+\bar{x}) = \frac{\alpha}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x}) + \frac{\alpha\bar{x}}{n}\sum_{i=1}^{n}(x_i-\bar{x}) \\ =&\alpha s_x^2\end{aligned}$$

であり，さらに，

$$\sum_{i=1}^{n} p_i^2\sigma^2 =\sigma^2\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n^2} = \frac{\sigma^2}{n}s_x^2.$$

演習 13　《例 5.1》のデータについて，

(1) α, β の 95% 信頼区間を作れ．

解答 $n=10$, $\bar{x}=170$, $s_{x^2}=6^2=36$,$\hat{\alpha}=0.5$, $\hat{\beta}=74$, $S_{ee}=225/2$ であった．したがって，$\hat{\sigma}^2=\frac{1}{10-2}\frac{225}{2}=\frac{225}{16}$, $\hat{\sigma}_\alpha^2=\frac{225/16}{36}=\frac{25}{64}$, $\hat{\sigma}_\beta^2=0\left(1+\frac{170^2}{36}\right)\frac{225}{16}=90425$ であり，また，$t_{10-2}(0.025)=2.306$ である．したがって，α の信頼区間は，$0.5\pm 2.306\times\frac{\sqrt{25/64}}{\sqrt{10}}=0.0442367\sim 0.955763$ であり，β の信頼区間は $74\pm 2.306\times\frac{\sqrt{90425}}{\sqrt{10}}=-145.282\sim 293.282$ となる．

(2)

$$\begin{cases} H_0 & : \ \alpha=0 \\ H_1 & : \ \alpha\neq 0\end{cases}$$

を有意水準 5% で検定しなさい．

解答 $S_{RR}=S_{yy}-S_{ee}=90$ だったので，$W=(10-2)\times\frac{90}{225/2}=\frac{32}{5}=6.4>f_{1,8}(0.05)=5.32$ となり，H_0 は棄却される．

演習 14　《例 6.1》のデータに関して，α_1, α_2 の最小 2 乗推定量を求めよ．また，自由度調整済み寄与率を求めよ．さらに，$H_0:\alpha_1=\alpha_2=0$, $H_1:\alpha_1\neq 0\vee\alpha_2\neq 0$ に対して仮説検定を行え．仮説検定に用いる諸量を分散分析表にまとめよ．

解答 $p=2$ なので，

$$\hat{\alpha}_1 = \frac{4/5 - 3/4 \cdot 3/5}{1-(3/5)^2} \cdot \frac{8}{7} = \frac{5}{8}, \quad \hat{\alpha}_2 = \frac{3/4 - 4/5 \cdot 3/5}{1-(3/5)^2} \frac{8}{7} = \frac{3}{8},$$

$$\begin{aligned} S_{ee} =& 12\left\{8^2 + \left(\frac{5}{8}\cdot 7\right)^2 + \left(\frac{3}{8}\cdot 9\right)^2 - 2\cdot\frac{5}{8}\cdot\frac{4}{5}\cdot 8\cdot 7 - 2\cdot\frac{3}{8}\cdot\frac{3}{4}\cdot 8\cdot 9\right. \\ &\left. +2\cdot\frac{5}{8}\cdot\frac{3}{8}\cdot\frac{3}{5}\cdot 7\cdot 9\right\} = 189. \end{aligned}$$

$S_{yy} = ns_{yy} = 12\cdot 8^2 = 768$, $n = 12$ なので，自由度調整済み寄与率は

$$R^{*2} = 1 - \frac{189/(12-2-1)}{768/(12-1)} = \frac{179}{256} = 0.699219.$$

さらに，$S_{RR} = S_{yy} - S_{ee} = 768 - 189 = 579$ なので，

$$W = \frac{S_R/p}{S_e/(n-p-1)} = \frac{579/2}{189/9} = \frac{193}{14} = 13.7857$$

$f_{p,n-p-1}(\alpha) = f_{2,9}(0.05) = 19.39$ だから，H_0 は棄却しない．

演習 15

$$S(a_1,a_2,b) = \sum_{i=1}^{n}\left(y_i - (a_1x_{i1} + a_2x_{i2} + b)\right)^2$$

の最小化について以下の問いに答えよ．ただし，$\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$, $s_{yy} = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2$, $\bar{x}_j = \frac{1}{n}\sum_{i=1}^n x_{ij}$, $s_{jj'} = \frac{1}{n}\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'})$, $s_{jy} = \frac{1}{n}\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})$, $r_{12} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}$, $r_{jy} = \frac{s_{jy}}{\sqrt{s_{jj}s_{yy}}}$, $Y_i = y_i - \bar{y}$, $X_{ij} = x_{ij} - \bar{x}_j$ とする．

(1) $T(a_1,a_2) = \sum_{i=1}^n \left(Y_i - (a_1X_{i1} + a_2X_{i2})\right)^2$ として，以下を示せ：

$$S(a_1,a_2,b) = T(a_1,a_2) + n(b - \bar{y} + a_1\bar{x}_1 + a_2\bar{x}_2)^2 \geqq T(a_1,a_2),$$

等号は $b = \bar{y} - a_1\bar{x}_1 - a_2\bar{x}_2$ の時のみ．

[ヒント：$\sum_{i=1}^n X_{ij} = 0$, $\sum_{i=1}^n Y_i = 0$ を使え]

解答

$$\begin{aligned} S(a_1,a_2,b) &= \sum_{i=1}^{n}(y_i - (a_1x_{i1} + a_2x_{i2} + b))^2 \\ &= \sum_{i=1}^{n}\left\{Y_i + \bar{y} - (a_1(X_{i1} + \bar{x}_1) + a_2(X_{i2} + \bar{x}_2) + b)\right\}^2 \end{aligned}$$

$$
\begin{aligned}
&=\sum_{i=1}^{n}\{Y_i-(a_1X_{i1}+a_2X_{i2})-(b-\bar{y}+a_1\bar{x}_1+a_2\bar{x}_2)\}^2\\
&=\sum_{i=1}^{n}(Y_i-(a_1X_{i1}+a_2X_{i2}))^2\\
&\qquad-2(b-\bar{y}+a_1\bar{x}_1+a_2\bar{x}_2)\sum_{i=1}^{n}(Y_i-(a_1X_{i1}+a_2X_{i2}))\\
&\qquad+n(b-\bar{y}+a_1\bar{x}_1+a_2\bar{x}_2)^2\\
&=T(a_1,a_2)+n(b-\bar{y}+a_1\bar{x}_1+a_2\bar{x}_2)^2\geqq T(a_1,a_2).
\end{aligned}
$$

ここで，$b=\bar{y}-a_1\bar{x}_1-a_2\bar{x}_2$ の時のみ，最後の不等号が等号となることは明らか.

(2) $a_1=\hat{\alpha}_1$, $a_2=\hat{\alpha}_2$ の時，$T(a_1,a_2)$ が最小になるとする．このとき，$\hat{\alpha}_1$, $\hat{\alpha}_2$ は

$$
\begin{cases}
s_{11}\hat{\alpha}_1+s_{12}\hat{\alpha}_2= & s_{1y}\\
s_{21}\hat{\alpha}_1+s_{22}\hat{\alpha}_2= & s_{2y}
\end{cases}
$$

を満たすことを示せ．これから，$\hat{\alpha}_1$, $\hat{\alpha}_2$ が次のように求められることを示せ.

$$
\hat{\alpha}_1=\frac{r_{1y}-r_{12}r_{2y}}{1-r_{12}^2}\sqrt{\frac{s_{yy}}{s_{11}}},\qquad \hat{\alpha}_2=\frac{r_{2y}-r_{12}r_{1y}}{1-r_{12}^2}\sqrt{\frac{s_{yy}}{s_{22}}}
$$

[ヒント:$j=1,2$ に対して，$\frac{\partial T}{\partial a_j}(\hat{\alpha}_1,\hat{\alpha}_2)=0$, $\sum_{i=1}^{n}X_{ij}X_{ij'}=ns_{jj'}$, $\sum_{i=1}^{n}X_{ij}Y_i=ns_{jy}$.]

解答 $a_1=\hat{\alpha}_1$, $a_2=\hat{\alpha}_2$ のとき，$T(a_1,a_2)$ は最小なので，

$$
\begin{aligned}
\frac{\partial T}{\partial a_1}(\hat{\alpha}_1,\hat{\alpha}_2)=&\sum_{i=1}^{n}2(Y_i-(\hat{\alpha}_1X_{i1}+\hat{\alpha}_2X_{i2})(-X_{i1})\\
=&-2\left(\sum_{i=1}^{n}Y_iX_{i1}-\hat{\alpha}_1\sum_{i=1}^{n}X_{i1}^2-\hat{\alpha}_2\sum_{i=1}^{n}X_{i1}X_{i2}\right)\\
=&-2n\left(s_{1y}-\hat{\alpha}_1s_{11}-\hat{\alpha}_2s_{12}\right)=0
\end{aligned}
$$

が成り立つ．これを簡単にすると，

$$
s_{11}\hat{\alpha}_1+s_{12}\hat{\alpha}_2=s_{1y}
$$

同様に，

$$
\begin{aligned}
\frac{\partial T}{\partial a_2}(\hat{\alpha}_1,\hat{\alpha}_2)=&\sum_{i=1}^{n}2(Y_i-(\hat{\alpha}_1X_{i1}+\hat{\alpha}_2X_{i2})(-X_{i2})\\
=&-2n\left(s_{2y}-\hat{\alpha}_1s_{21}-\hat{\alpha}_2s_{22}\right)=0
\end{aligned}
$$

から，

$$s_{21}\hat{\alpha}_1 + s_{22}\hat{\alpha}_2 = s_{2y}$$

が導ける．

これらを連立して解くと，

$$\begin{aligned}\hat{\alpha}_1 =& \frac{s_{22}s_{1y} - s_{12}s_{2y}}{s_{22}s_{11} - s_{12}s_{21}} \\ =& \frac{\frac{s_{1y}}{s_{11}} - \frac{s_{12}s_{2y}}{s_{11}s_{22}}}{1 - \frac{s_{12}s_{21}}{s_{11}s_{22}}} = \frac{\left(\frac{s_{1y}}{\sqrt{s_{11}s_{yy}}} - \frac{s_{12}}{\sqrt{s_{11}s_{22}}}\frac{s_{2y}}{\sqrt{s_{22}s_{yy}}}\right)\frac{\sqrt{s_{yy}}}{\sqrt{s_{11}}}}{1 - \left(\frac{s_{12}}{\sqrt{s_{11}s_{22}}}\right)^2} \\ =& \frac{r_{1y} - r_{12}r_{2y}}{1 - r_{12}^2}\sqrt{\frac{s_{yy}}{s_{11}}}\end{aligned}$$

であり，

$$\begin{aligned}\hat{\alpha}_2 =& \frac{s_{21}s_{1y} - s_{11}s_{2y}}{s_{21}s_{12} - s_{11}s_{22}} \\ =& \frac{r_{2y} - r_{12}r_{1y}}{1 - r_{12}^2}\sqrt{\frac{s_{yy}}{s_{22}}}\end{aligned}$$

【注意】$\hat{\beta} = \bar{y} - \hat{\alpha}_1\bar{x}_1 - \hat{\alpha}_2\bar{x}_2$ とすると，(1) より，

$$S(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}) = T(\hat{\alpha}_1, \hat{\alpha}_2) + n(\hat{\beta} - \bar{y} + \hat{\alpha}_1\bar{x}_1 + \hat{\alpha}_2\bar{x}_2)^2 = T(\hat{\alpha}_1, \hat{\alpha}_2)$$

となる．また，$a_1 = \hat{\alpha}_1, a_2 = \hat{\alpha}_2$ のとき，$T(a_1, a_2)$ は最小なので，任意の a_1, a_2 に対して，$T(a_1, a_2) \geqq T(\hat{\alpha}_1, \hat{\alpha}_2)$ となる．したがって，任意の a_1, a_2, b に対して，

$$S(a_1, a_2, b) \geqq T(a_1, a_2) \geqq T(\hat{\alpha}_1, \hat{\alpha}_2) = S(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}).$$

つまり，$a_1 = \hat{\alpha}_1$, $a_2 = \hat{\alpha}_2$, $b = \hat{\beta}$ の時，$S(a_1, a_2, b)$ は最小値をとることがわかる．

(3) $\hat{\beta} = \bar{y} - \hat{\alpha}_1\bar{x}_1 - \hat{\alpha}_2\bar{x}_2$, $\hat{y}_i = \hat{\alpha}_1 x_{i1} + \hat{\alpha}_2 x_{i2} + \hat{\beta}$, $\hat{e}_i = y_i - \hat{y}_i$ とすると，

$$\hat{y}_i - \bar{y} = \hat{\alpha}_1 X_{i1} + \hat{\alpha}_2 X_{i2}, \quad \hat{e}_i = Y_i - (\hat{\alpha}_1 X_{i1} + \hat{\alpha}_2 X_{i2})$$

が成り立つことを示せ．さらに，

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})\hat{e}_i = 0$$

を示せ．

[ヒント：$j = 1, 2$ に対して，$\frac{\partial T}{\partial a_j}(\hat{\alpha}_1, \hat{\alpha}_2) = 0.$]

解答

$$
\begin{aligned}
\hat{y}_i - \bar{y} =& \hat{\alpha}_1 x_{i1} + \hat{\alpha}_2 x_{i2} + \hat{\beta} - \bar{y} \\
=& \hat{\alpha}_1 x_{i1} + \hat{\alpha}_2 x_{i2} + \bar{y} - \hat{\alpha}_1 \bar{x}_1 - \hat{\alpha}_2 \bar{x}_2 - \bar{y} \\
=& \hat{\alpha}_1 (x_{i1} - \bar{x}_1) + \hat{\alpha}_2 (x_{i2} - \bar{x}_2) = \hat{\alpha}_1 X_{i1} + \hat{\alpha}_2 X_{i2}, \\
\hat{e}_i =& y_i - \hat{y}_i = y_i - \bar{y} - (\hat{y}_i - \bar{y}) \\
=& Y_i - (\hat{\alpha}_1 X_{i1} + \hat{\alpha}_2 X_{i2}).
\end{aligned}
$$

さらに，(2) において，

$$
\frac{\partial T}{\partial a_1} T(\hat{\alpha}_1, \hat{\alpha}_2) = \sum_{i=1}^{n} 2(Y_i - (\hat{\alpha}_1 X_{i1} + \hat{\alpha}_2 X_{i2})(-X_{i1}) = 0
$$

だったので，

$$
\sum_{i=1}^{n} \hat{e}_i X_{i1} = 0
$$

が成り立つ．同様に，

$$
\frac{\partial T}{\partial a_1} T(\hat{\alpha}_1, \hat{\alpha}_2) = \sum_{i=1}^{n} 2(Y_i - (\hat{\alpha}_1 X_{i1} + \hat{\alpha}_2 X_{i2})(-X_{i2}) = 0
$$

だったので，

$$
\sum_{i=1}^{n} \hat{e}_i X_{i2} = 0
$$

が成り立つ．したがって，

$$
\begin{aligned}
\sum_{i=1}^{n} (\hat{y}_i - \bar{y}) \hat{e}_i =& \sum_{i=1}^{n} \hat{e}_i (\hat{\alpha}_1 X_{i1} + \hat{\alpha}_2 X_{i2}) \\
=& \hat{\alpha}_1 \sum_{i=1}^{n} \hat{e}_i X_{i1} + \hat{\alpha}_2 \sum_{i=1}^{n} \hat{e}_i X_{i2} = 0
\end{aligned}
$$

がわかる．

演習 16　ある地域の山岳地帯における樹木について，その年輪 (y) と太さ (x inch, 1inch=2.54cm) を測定した．その結果，年輪の平均，標準偏差は 106.42, 49.5, 太さの平均，標準偏差は 24.36, 12.82 であり，相関係数は 0.8279 であった．ただし，測定した樹木の本数は 24 本である．

(1) 年輪の太さに対する回帰直線 $y = ax + b$ を最小 2 乗法で推定せよ．

解答　$s_{xy} = 0.8279 \times 49.5 \times 12.82 = 525.377$, $s_x^2 = 12.82^2 = 164.352$ なので，$\hat{a} = \frac{0.8279 \times 49.5 \times 12.82}{12.82^2} = 3.197$. $\hat{b} = 106.42 - 3.197 \times 24.36 = 28.55$ なので，$y = 3.197x + 28.55$.

(2) 回帰による寄与率を求めよ．

解答　寄与率 R^2 は相関係数の 2 乗なので，$R^2 = 0.685418$.

(3) 仮説 $H_0 : a = 0$, $H_1 : a \neq 0$ を有意水準 5% で検定せよ．

解答　$S_e = 24\left(49.5^2 - \frac{525.377^2}{12.82^2}\right) = 18499.3$, $\hat{\sigma}^2 = \frac{18499.3}{22} = 840.877$, $\hat{\sigma}_a^2 = \frac{840.877}{12.82^2} = 5.1163$．よって，$T_a^2 = \frac{n\hat{a}^2}{\sigma_a^2} = \frac{24\times 3.197^2}{5.1163} = 47.9342$. $t_{24-2}(0.025) = 2.074^2 = 4.30148$ だから，$|T_a| > t_{22}(0.05)$ であり，H_0 が棄却される．つまり，a は有意に 0 でない．

(4) 太さが 30 inch の樹木について，年輪の平均値の 95% 両側信頼区間を求めよ．

解答　$x_0 = 30$ の時の y の値 $y_0 = ax_0 + b$ の推定値は $\hat{y}_0 = \hat{a}x_0 + \hat{b} = 124.45$ であり，その 95% 両側信頼区間は，

$$y_0 \in 124.45 \pm 2.074\frac{\sqrt{840.877}}{\sqrt{24}}\sqrt{1 + \frac{(30-24.36)^2}{12.82^2}} = [77.99, 170.91]\,.$$

(5) 太さが 30 inch の樹木を選んで新たに年輪を測定する時，その測定値を信頼度 95% で予測せよ．

解答　$x_{25} = 30$ の時の y_{25} の値 $y_{25} = ax_{25} + b + \epsilon_{25}$ の予測値は $\hat{y}_{25} = \hat{a}x_{25} + \hat{b} = 124.45$ であり，その 95% 予測区間は，

$$y_0 \in 124.45 \pm 2.074\frac{\sqrt{840.877}}{\sqrt{24}}\sqrt{24 + 1 + \frac{(30-24.36)^2}{12.82^2}} = [-89.0, 337.90]\,.$$

演習 17　日経平均株価 y と内閣支持率 x_1, 失業率 x_2 の月間データを 22ヶ月調査したところ，つぎのような結果を得た．

$$s_y = 900,\quad s_{x_1} = 1/3,\quad s_{x_2} = 1/6,\quad r_{yx_1} = 2/3,\quad r_{yx_2} = -3/4,\quad r_{x_1x_2} = -4/5.$$

回帰モデル $y_i = a_1x_{i1} + a_2x_{i2} + b + \varepsilon_i$ を仮定して，つぎの問に答えよ．

(1) 最小 2 乗推定量 $\hat{a}_1$, $\hat{a}_2$ を求めよ．

解答　$\hat{a}_1 = \frac{r_{yx_1} - r_{x_1x_2}r_{yx_2}}{1 - r_{x_1x_2}^2}\sqrt{\frac{s_y^2}{s_{x_1}^2}} = \frac{2/3-(-4/5)\cdot(-3/4)}{1-(-4/5)^2}\sqrt{\frac{900^2}{(1/3)^2}} = 500$.　$\hat{a}_2 = \frac{r_{2y} - r_{12}r_{1y}}{1 - r_{12}^2}\sqrt{\frac{s_y^2}{s_{x_2}^2}} = \frac{-3/4-(-4/5)\cdot(2/3)}{1-(-4/5)^2}\sqrt{\frac{900^2}{(1/6)^2}} = -3250$.

(2) y と x_1, x_2 の重相関係数を求めよ．

解答　$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}} = \sqrt{\frac{(2/3)^2+(-3/4)^2-2\cdot(2/3)\cdot(-3/4)\cdot(-4/5)}{1-(-4/5)^2}} = \frac{\sqrt{745}}{36} = 0.758$

(3) x_2 の影響を除いた y と x_1 の偏相関係数を求め，r_{yx_1} を比較せよ．

解答　$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{(1-r_{yx_2}^2)(1-r_{x_1x_2}^2)}} = \frac{2/3-(-3/4)\cdot(-4/5)}{\sqrt{(1-(-3/4)^2)(1-(-4/5)^2)}} = \frac{4}{9\sqrt{7}} = 0.167984 < 2/3 = r_{yx_1}$ であり，日経平均株価と内閣支持率の関係の大部分は失業率に依存するものであり，直接の関係はそれほど大きくない．

(4) $H_0\ :\ a_1 = a_2 = 0$, $H_1\ :\ H_0$ でない を有意水準 5% で検定せよ．

解答

$$S_e = \frac{n(1 - r_{yx_2}^2 - r_{x_1x_2}^2 - r_{yx_1}^2 + 2r_{x_1x_2}r_{yx_1}r_{yx_2})}{1 - r_{x_1x_2}^2}s_y^2$$

に代入すると，$S_e = 7576250$ がわかる．$S_y = ns_y^2 = 17820000$ だから，$S_R = S_y - S_e = 10243750$. したがって，$W = \frac{S_r/2}{S_e/(22-2-1)} = \frac{745}{58} = 12.8448$. $f_{2,19}(0.05) = 3.52$ だから，H_0 は棄却される．つまり，日経平均株価は内閣支持率および失業率によって説明できる．

演習 18　日本人の血液型は O 型が 35%, A 型が 30%, B 型が 25%, AB 型が 10% である．政治家 200 人を調べたところ O 型が 83 人，A 型が 45 人，B 型が 62 人，AB 型が 10 人であった．政治家の比率は一般の日本人と異なっているとしてよいか有意水準 5% で検定せよ．

解答 $p_1 = 0.35, p_2 = 0.3, p_3 = 0.25, p_4 = 0.1, x_1 = 83, x_2 = 45, x_3 = 62, x_4 = 10,$ $n = x_1 + \cdots + x_4 = 200$

$$\begin{aligned} V =& \frac{(83 - 200 \times 0.35)^2}{200 \times 0.35} + \frac{(45 - 200 \times 0.3)^2}{200 \times 0.3} + \frac{(62 - 200 \times 0.25)^2}{200 \times 0.25} \\ &+ \frac{(10 - 200 \times 0.1)^2}{200 \times 0.1} = \frac{9831}{700} = 14.04 > \chi_3^2(0.05) = 7.815 \end{aligned}$$

だから，一般の日本人と有意に異なる．

演習 19 あるコンビニにおけるおにぎりの販売個数を，陳列する場所 A, B, C と時間帯に分けて集計したところ，以下のようになった．陳列する場所と時間帯は独立かどうか有意水準 5% で仮説検定せよ．

	A	B	C
午前	20	32	48
午後	20	50	30

解答 $x_{11} = 20,\ x_{21} = 20,\ x_{12} = 32,\ x_{22} = 50,\ x_{13} = 48,\ x_{23} = 30,\ n = x_{11} + x_{21} + \cdots + x_{23} = 200,\ x_{1\cdot} = 40,\ x_{2\cdot} = 82,\ x_{3\cdot} = 78,\ x_{\cdot 1} = 100,\ x_{\cdot 2} = 200.$ $n\hat{p}_{11} = 20, n\hat{p}_{12} = 41, n\hat{p}_{13} = 39, n\hat{p}_{21} = 20, n\hat{p}_{22} = 41, n\hat{p}_{23} = 39$ なので，

$$\begin{aligned} V =& \frac{(20-20)^2}{20} + \frac{(32-41)^2}{41} + \frac{(48-39)^2}{39} \\ & \frac{(20-20)^2}{20} + \frac{(50-41)^2}{41} + \frac{(30-39)^2}{39} = \frac{4320}{533} = 8.10507 > \chi_2^2(0.05) = 5.991 \end{aligned}$$

だから，陳列位置と時間帯は独立とは言えない．

演習 20　数学に関するアンケートをとったところ，好きが 48 人，嫌いが 32 人，その他が 20 人であった．好きが嫌いより多いといえるかどうか有意水準 5% と 1% で検定せよ．

解答　　母集団における好きな人の割合を p_1, 嫌いな人の割合を p_2 とする．n 人にアンケートをとって，好きな人が X, 嫌いな人が Y 人の時，$\bar{X} = X/n$, $\bar{Y} = Y/n$ とすると，

$$Z = \sqrt{n}\frac{\bar{X} - \bar{Y} - (p_1 - p_2)}{\sqrt{(\bar{X} + \bar{Y}) - (\bar{X} - \bar{Y})^2}}$$

は近似的に正規分布 $N(0, 1)$ に従う．今，帰無仮説 $H_0 \ : \ p_1 = p_2$, 対立仮説 $H_1 \ : \ p_1 > p_2$ とすると，$H_1 \ : \ p_1 > p_2 \iff p_1 - p_2 > 0$ であり，$p_1 \approx \bar{X}$, $p_2 \approx \bar{Y}$ だから，棄却域 R の形は

$$R = \{\bar{X} - \bar{Y} > A\}$$

となる．帰無仮説が正しい時，$p_1 = p_2$ であり，そのとき，

$$Z = \sqrt{n}\frac{\bar{X} - \bar{Y}}{\sqrt{(\bar{X} + \bar{Y}) - (\bar{X} - \bar{Y})^2}} =: Z_0$$

が近似的に $N(0, 1)$ に従うことを考えると，帰無仮説が正しい時，帰無仮説を棄却する確率 α は，

$$\alpha = P\{\bar{X} - \bar{Y} > A\} = P\left\{Z_0 > \frac{\sqrt{n}A}{\sqrt{(\bar{X} + \bar{Y}) - (\bar{X} - \bar{Y})^2}}\right\}$$

と表せるので，

$$\frac{\sqrt{n}A}{\sqrt{(\bar{X} + \bar{Y}) - (\bar{X} - \bar{Y})^2}} = z(\alpha)$$

となる．つまり，$R = \{Z_0 > z(\alpha)\}$ である．

さて，アンケートの結果，$X = 48$, $Y = 32$, $n = 100$, $\bar{X} = X/n = 0.48$, $\bar{Y} = Y/n = 0.32$ だから，$Z_0 = \frac{20}{11} = 1.81818$. $z(0.05) = 1.645$, $z(0.01) = 2.32666$ だから，有意水準 5% では好きが嫌いより有意に多いと言えるが，有意水準 1% では有意に好きが多いと言えない．