第 10 章

多標本問題

第 6 章の 2 標本の 2 を 3 以上に拡張したモデルで 3 種の手法のよさの関係も 2 標本問題の場合と同じであるが, 手法の構成法と分布論を学習する必要があり, 多重比較と呼ばれる問題への自然な理論を作ることができる.

10.1 モデル

ある要因 A があり, k 個の水準 $A_1, \cdots, A_k$ を考え, 水準 A_i における標本 $(X_{i1}, X_{i2}, \cdots, X_{in_i})$ は第 i 標本と呼ばれ, 各成分は同一の連続型分布関数 $F(\dfrac{x-\mu_i}{\sigma})$ をもつとする. さらにすべての X_{ij} $(j=1,\cdots,n_i; i=1,\cdots,k)$ は互いに独立であると仮定する. 一般性を失うことなく $\displaystyle\int_{-\infty}^{\infty} xdF(x) = 0$, $\displaystyle\int_{-\infty}^{\infty} x^2 dF(x) = 1$ とする. このとき, $E(X_{ij}) = \mu_i$, $V(X_{ij}) = \sigma^2$ である. 総標本数を $n \equiv n_1 + \cdots + n_k$ とおき,

$$\nu \equiv \frac{1}{n}\sum_{i=1}^{k} n_i\mu_i, \qquad \alpha_i \equiv \mu_i - \nu$$

とおく, このとき $\displaystyle\sum_{i=1}^{k} n_i\alpha_i = 0$ が成り立つ. α_i は要因 A の水準 A_i における**主効果**, ν を**全平均**とよぶ. ここで

$$X_{ij} = \mu_i + \varepsilon_{ij} = \nu + \alpha_i + \varepsilon_{ij},\ \text{ただし}\ \sum_{i=1}^{k} n_i\alpha_i = 0$$

と書き直せ, ε_{ij} は誤差確率変数とよばれ独立で同一の分布関数 $F(\dfrac{x}{\sigma})$ をもつ. 仮説検定として,

帰無仮説 $H_0:\ \mu_1 = \cdots = \mu_k$ と 対立仮説 $H_1:$ ある $i \neq i'$ について $\mu_i \neq \mu_{i'}$

に興味がある. また, この帰無仮説と対立仮説は

$$H_0:\ \alpha_1 = \cdots = \alpha_k = 0\ \text{と}\ H_1:\ \text{ある}\ i\ \text{について}\ \alpha_i \neq 0$$

と同値である. また, パラメータ $(\alpha_1, \cdots, \alpha_k)$, ν, $(\mu_1, \cdots, \mu_k)$ の点推定に興味がある.

10.2　正規性の検定と分布の探索

第 i 標本の標本平均を, $\bar{X}_{i\cdot} \equiv \dfrac{1}{n_i}\sum_{j=1}^{n_j} X_{ij}$ とおく. $\sum_{i'=0}^{i-1} n_{i'} + 1 \leq m \leq \sum_{i'=0}^{i} n_{i'}$ となる整数 m と $j \equiv m - \sum_{i'=0}^{i-1} n_{i'}$ に対して

$$Z_m \equiv (X_{ij} - \bar{X}_{i\cdot})\Big/\sqrt{1 - \frac{1}{n_i}} \qquad (j = 1, \cdots, n_i) \tag{10.1}$$

で $Z_1, \cdots, Z_n$ を定義する, ただし, $n_0 \equiv 0$ とおく. このとき $E(Z_i) = 0$, $V(Z_i) = \sigma^2$ である. $\hat{G}_n(x)$ を

$$\hat{G}_n(x) \equiv \frac{1}{n}\#\{Z_i : Z_i \leq x, 1 \leq i \leq n\} = \frac{1}{n}\{x\ \text{以下となる}\ Z_i\text{の個数}\ \}$$

で定義する.

観測値の従っている分布を調べる方法として,

$$D_F \equiv \sup_{-\infty < x < \infty} |\hat{G}_n(x) - F(\frac{x}{\breve{\sigma}_n})| \tag{10.2}$$

とおく, ただし,

$$\breve{\sigma}_n \equiv \frac{1}{n\int_{-\infty}^{\infty}|x|dF(x)}\sum_{i=1}^{n}|Z_i|.$$

$\breve{\sigma}_n$ は σ の頑健推定量である．

[1] **正規性の検定**; 『帰無仮説 $H: F(x)$ は正規分布 v.s. 対立仮説 $H^c: F(x)$ は正規分布ではない』の検定として

$$D_\Phi = \sup_{-\infty<x<\infty}|\hat{G}_n(x) - \Phi(\frac{x}{\breve{\sigma}_n})| \tag{10.3}$$

が大きいとき H を棄却する方法が採用できる．シミュレーションによる検定手順を以下に示す．

- 手順 1. $0 < \alpha < 1$ となるように有意水準 α を決め, D_Φ を計算する．
- 手順 2. シミュレーションの繰り返し数 K として 1000 を採り,
  $\ell = 1, 2, \cdots, K-1$ について以下の手順 2.1 と手順 2.2 を繰り返す．
  - · 手順 2.1. 正規乱数を n 個生成させ, それを $\{X_{ij}(\ell):\ j = 1, \cdots, n_i,$
    $i = 1, \cdots, k\}$ とし, (1) と同様に $Z_1(\ell), \cdots, Z_n(\ell)$ を定義する．
  - · 手順 2.2. $Z_1, \cdots, Z_n$ の替わりに $Z_1(\ell), \cdots, Z_n(\ell)$ で (10.3) 式を計算し,
    それを $D_\Phi(\ell)$ とする．
- 手順 3. $N \equiv \sum_{\ell=1}^{K-1} I(D_\Phi < D_\Phi(\ell))$ とおく．$\dfrac{N}{K}$ は p 値となり, この値が
  α より小さいならば H を棄却する．

残念ながら, 1 標本の場合と同様に観測値の従っている分布が正規分布でないときでも n が非常に大きくないと正規性を棄却できないことが多い．

[2] **分布の探索**; D_F は, 観測値の従っている分布がどれくらい $F(\dfrac{x-\mu}{\sigma})$ に近いかのはかりとして見れる．$F(x)$ として正規分布, 混合正規分布, ロジスティック分布, 両側指数分布, 異常値を持つ混合正規分布, ワイブル分布, 対数正規分布, 指数分布を当てはめ, 統計量 D_F のもっとも小さい値を与える分布 $F(x)$ を探し, それを観測値の従っている分布に最も近い分布と見なすことができる．

10.3　正規母集団での最良手法

$F(x)$ を標準正規分布 $N(0,1)$ の分布関数とする．すなわち，$F(\frac{x-\mu_i}{\sigma})$ は正規分布 $N(\mu_i, \sigma^2)$ の分布関数．

[3] **F 検定**; $\bar{X}_{i\cdot} \equiv \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$, $\bar{X}_{\cdot\cdot} \equiv \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}$ とおくとき，

$$\begin{aligned} SSW &\equiv \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\cdot})^2 \\ SST &\equiv \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{\cdot\cdot})^2 \end{aligned}$$

をそれぞれ，**級内変動** (Sum of Squares Within)，**総変動** (Sum of Squares Total) といい，

$$SSB \equiv SST - SSW = \sum_{i=1}^{k} n_i(\bar{X}_{i\cdot}-\bar{X}_{\cdot\cdot})^2 \tag{10.4}$$

の関係がある．この (10.4) を**級間変動** (Sum of Squares Between) という．これらを使って帰無仮説 H_0 と対立仮説 H_1 の検定統計量は

$$T_S = \frac{SSB/(k-1)}{SSW/(n-k)}$$

と表され，T_S が大きいとき H_0 を棄却する．

定理 10.1　H_0 の下で，T_S は自由度 $(k-1, n-k)$ の F 分布に従う．

証明　一般性を失うことなく $\mu_i = 0, \sigma^2 = 1$ とする．系 3.8 より $\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2 \sim \chi^2_{n_i-1}$．これにより $SSW \sim \chi^2_{n-k}$．また，$SSB = \sum_{i=1}^{k}\frac{n_i}{n}\{\sqrt{n}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})\}^2$．ここで定理 3.8 を適用すると，$SSB \sim \chi^2_{k-1}$．
$\boldsymbol{Y} \equiv {}^t(X_{11}-\bar{X}_{1\cdot}, \cdots, X_{1n_1-1}-\bar{X}_{1\cdot}, X_{21}-\bar{X}_{2\cdot}, \cdots, X_{2n_2-1}-\bar{X}_{2\cdot}, \cdots,$
$X_{kn_k-1}-\bar{X}_{k\cdot})$, $\boldsymbol{Z} \equiv {}^t(\bar{X}_{1\cdot}-\bar{X}_{\cdot\cdot}, \cdots, \bar{X}_{k-1\cdot}-\bar{X}_{\cdot\cdot})$ とおくと ${}^t({}^t\boldsymbol{Y}, {}^t\boldsymbol{Z})$ は $n-1$ 次元正規分布に従い，$\mathrm{Cov}(X_{ij}-\bar{X}_{i\cdot}, \bar{X}_{m\cdot}-\bar{X}_{\cdot\cdot}) = 0$ となる．命題 3.2

より, $\boldsymbol{Y}$ と $\boldsymbol{Z}$ は互いに独立である. 定理 B.2 より $\boldsymbol{Y}$ の関数である SSW と $\boldsymbol{Z}$ の関数である SSB はたがいに独立である. これにより結果を得る. □

$F^{k-1}_{n-k,\alpha}$ を自由度 $(k-1, n-k)$ の F 分布の上側 α 点とすると, 定理 10.1 により, $P_0(T_S > F^{k-1}_{n-k,\alpha}) = \alpha$. ゆえに有意水準 α の検定方式は,

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & (T_S > F^{k-1}_{n-k,\alpha} \text{のとき}) \\ 0 & (T_S < F^{k-1}_{n-k,\alpha} \text{のとき}) \end{cases}$$

と表現される. (α と α_i は別のもの)

[4] **点推定量**; パラメータ α_i, ν, μ_i の推定量は

$$\tilde{\alpha}_i = \bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}\ (i = 1, \cdots, k),\ \tilde{\nu} = \bar{X}_{\cdot\cdot},\ \tilde{\mu}_i = \bar{X}_{i\cdot}\ (i = 1, \cdots, k).$$

10.4 ノンパラメトリック法

分布関数 $F(x)$ は未知でもかまわないとする.

[5] **クラスカル • ワリスの順位検定**;

3.4 節 (10) で定義したように n 個すべての観測値 $X_{11}, \cdots, X_{kn_k}$ を小さい方から並べたときの X_{ij} の順位を R_{ij} とする. $\bar{R}_{i\cdot} \equiv \dfrac{1}{n_i}\sum_{j=1}^{n_i} R_{ij}$ とし

$$T_R \equiv \sum_{i=1}^{k} n_i(\bar{R}_{i\cdot} - \frac{n+1}{2})^2$$

とおくと, H_0 の下で (3.5) が成り立ち, T_R の分布は

$$P(T_R \le t) = \frac{1}{n!}\#\{\boldsymbol{r} : \sum_{i=1}^{k} n_i(\bar{r}_{i\cdot} - \frac{n+1}{2})^2 \le t, \boldsymbol{r} \in \mathcal{R}_n\}$$

である, ただし, $\boldsymbol{r} \equiv (r_{11}, \cdots, r_{kn_k})$, $\bar{r}_{i\cdot} \equiv \dfrac{1}{n_i}\sum_{j=1}^{n_i} r_{ij}$, $\mathcal{R}_n$ は (3.4) で定義したものとする. また, $S_i \equiv \sqrt{\dfrac{12}{n+1}}(\bar{R}_{i\cdot} - \dfrac{n+1}{2})$ $(i = 1, \cdots, k)$ とおき,

$$Z_R \equiv \frac{12}{n(n+1)} T_R = \sum_{i=1}^{k} (\frac{n_i}{n}) S_i^2$$

とおく．このとき, H_0 の下で $\boldsymbol{S} \equiv {}^t(S_1, \cdots, S_k)$ の平均と分散共分散行列はそれぞれ $E_0(\boldsymbol{S}) = \boldsymbol{0}$,

$$V_0(\boldsymbol{S}) = \mathrm{diag}(\frac{n}{n_1}, \cdots, \frac{n}{n_k}) - \boldsymbol{1}_k \cdot {}^t\boldsymbol{1}_k$$

となる (演習 10.6), ただし $\boldsymbol{1}_k = {}^t(\overbrace{1, \cdots, 1}^{k\text{ 個}})$.

定義 10.1　k 次正方行列 $\boldsymbol{A}$ に対して $\boldsymbol{AXA} = \boldsymbol{A}$ を満たす k 次正方行列 $\boldsymbol{X}$ を $\boldsymbol{A}$ の**一般化逆行列**といい, $\boldsymbol{A}^-$ で表す．　□

$\boldsymbol{A}$ が正則行列ならばその一般化逆行列は $\boldsymbol{A}^- = \boldsymbol{A}^{-1}$ となり一意である．しかしながら $\boldsymbol{A}$ が正則行列でないならばその一般化逆行列は 2 つ以上存在する．

定理 10.2　$Z_R = {}^t\boldsymbol{S}\{V_0(\boldsymbol{S})\}^- \boldsymbol{S}$ が成り立つ．

証明　$V_0(\boldsymbol{S})$ の階数は $k-1$. また $(n_1, \cdots, n_k)(V_0(\boldsymbol{S}), \boldsymbol{S}) = \boldsymbol{0}$ より, 行列 $(V_0(\boldsymbol{S}), \boldsymbol{S})$ の階数も $k-1$ である．ゆえに, $\boldsymbol{S} = V_0(\boldsymbol{S})\boldsymbol{x}$ となる列ベクトル $\boldsymbol{x}$ が存在する．これにより,

$${}^t\boldsymbol{S}\{V_0(\boldsymbol{S})\}^- \boldsymbol{S} = {}^t\boldsymbol{x} V_0(\boldsymbol{S})\{V_0(\boldsymbol{S})\}^- V_0(\boldsymbol{S})\boldsymbol{x} = {}^t\boldsymbol{x} V_0(\boldsymbol{S})\boldsymbol{x}$$

となり, この値は $V_0(\boldsymbol{S})$ の一般化逆行列の採り方に依らない．そこで
$\{V_0(\boldsymbol{S})\}^- = \mathrm{diag}(\frac{n_1}{n}, \cdots, \frac{n_k}{n})$ とすれば, ${}^t\boldsymbol{S}\{V_0(\boldsymbol{S})\}^- \boldsymbol{S} = Z_R$ を得る．　□

つぎの定理はよく知られていることであるが完全な証明にはかなりの知識を要するので, 証明は概略を述べる．

定理 10.3　$\displaystyle\lim_{n\to\infty} \frac{n_i}{n} = \lambda_i > 0$ $(i = 1, \cdots, k)$ とする．このとき, $n \to \infty$ として, H_0 の下で $Z_R \xrightarrow{\mathcal{L}} Z \sim \chi^2_{k-1}$, すなわち, Z_R は自由度 $k-1$ のカイ二乗分布に分布収束する．

証明　ハエック • シダック • セン (Hájek, Šidák and Sen(1999)) より,

$$\boldsymbol{S} \xrightarrow{\mathcal{L}} {}^t(Y_1 - \sum_{j=1}^{k} \lambda_j Y_j, \cdots, Y_k - \sum_{j=1}^{k} \lambda_j Y_j)$$

ただし, $Y_1, \cdots, Y_k$ は互いに独立で $Y_i \sim N(0, \dfrac{1}{\lambda_i})$. ここで定理 B.9 より $Z_R \xrightarrow{\mathcal{L}} \sum_{i=1}^{k} \lambda_i (Y_i - \sum_{j=1}^{k} \lambda_j Y_j)^2$. 定理 3.8 を適用して結果を得る. □

$T(\boldsymbol{X}) = T_R$ が大きいとき H_0 を棄却する. $0 < \alpha < 1$ に対して $P_0(T_R > t_\alpha) \leq \alpha$ かつ $P_0(T_R \geq t_\alpha) > \alpha$ となる t_α を探す. このとき, 有意水準 α の検定方式は検定関数 $\phi(\cdot)$ を使って,

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & (T_R > t_\alpha \text{のとき}) \\ \dfrac{\alpha - P_0(T_R > t_\alpha)}{P_0(T_R = t_\alpha)} & (T_R = t_\alpha \text{のとき}) \\ 0 & (T_R < t_\alpha \text{のとき}) \end{cases}$$

と表現される. $n_1, \cdots, n_k$ が大きいとき, カイ二乗分布の上側 α 点を $\chi^2_{k-1,\alpha}$ とおけば, 定理 10.3 より有意水準 α の検定方式は単純で検定関数 $\phi(\cdot)$ を使って,

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & (Z_R > \chi^2_{k-1,\alpha} \text{のとき}) \\ 0 & (Z_R < \chi^2_{k-1,\alpha} \text{のとき}) \end{cases}$$

と表現される.

[6] 順位推定量; $1 \leq i \neq i' \leq k$ に対して

$$\begin{aligned} \hat{\delta}_{ii'} &= \mathrm{med}\{X_{ij} - X_{i'j'} : j = 1, \cdots, n_i, j' = 1, \cdots, n_{i'}\} \\ &= (n_i \cdot n_{i'} \text{個の値 } \{X_{ij} - X_{i'j'} : j = 1, \cdots, n_i, j' = 1, \cdots, n_{i'}\} \text{ の中央値}) \end{aligned}$$

は 6.4 節 [7] より $\mu_i - \mu_{i'}$ のホッジス ● レーマン順位推定量である. また, $\alpha_i = \dfrac{1}{n} \sum_{i'=1}^{k} n_{i'}(\mu_i - \mu_{i'})$ より, α_i の順位推定量として

$$\hat{\alpha}_i = \frac{1}{n} \sum_{i'=1}^{k} n_{i'} \hat{\delta}_{ii'} \ (i = 1, \cdots, k)$$

を提案できる, ただし $\hat{\delta}_{ii} = 0$ とする. $1 \leq i \leq k, 1 \leq j \leq n_i$ に対して $Y_m \equiv X_{ij} - \hat{\alpha}_i$ とおく, ただし $m \equiv \sum_{p=1}^{i-1} n_p + j$, $\sum_{p=1}^{0} n_p = 0$ とする. このとき, ν の

順位推定量は

$$\begin{aligned}\hat{\nu} &= \mathrm{med}\{\frac{Y_\ell + Y_m}{2} : 1 \leq \ell \leq m \leq n\} \\ &= (\frac{n(n+1)}{2}\text{個の値 } \{\frac{Y_\ell + Y_m}{2} : 1 \leq \ell \leq m \leq n\} \text{ の中央値})\end{aligned}$$

であり, μ_i の順位推定量は

$$\hat{\mu}_i = \hat{\nu} + \hat{\alpha}_i \quad (i = 1, \cdots, k).$$

である．推定量の理論的正当性として, $\hat{\alpha}_i$ は密度関数 $f(x) \equiv F'(x)$ に対称性を必要としないが, $\hat{\nu}$, $\hat{\mu}_i$ には $f(x)$ の対称性を必要とする．

密度関数 $f(x)$ に対称性を仮定できない場合は, $\tilde{\nu} = \bar{X}_{\cdot\cdot}$ を ν の推定量とする．μ_i の推定量は

$$\hat{\mu}_i = \tilde{\nu} + \hat{\alpha}_i \quad (i = 1, \cdots, k).$$

10.5　セミパラメトリック法

前節と同様に, $F(x)$ は未知でもかまわないとする．

[7] **M 検定**; 帰無仮説 H_0 と 対立仮説 H_1 の有意水準 α の検定を考える．3.4 節 (11) で定義したように $X_{11}, \cdots, X_{kn_k}$ の順序統計量を $X_{(1)} \leq \cdots \leq X_{(n)}$ とおき, その実現値を $x_{(1)} \leq \cdots \leq x_{(n)}$ とする．関数 $\psi(\cdot)$ を $\psi(x) = \max(\min(x, b), -b)$ で定義し, [4] の $\tilde{\nu}$ を使って, 尺度母数を $\hat{\sigma}_n \equiv \frac{\sqrt{\pi}}{\sqrt{2} \cdot n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} |X_{ij} - \tilde{\nu}|$ で推定し, $\bar{\psi}_{i\cdot}(\boldsymbol{X}) \equiv \frac{1}{n_i} \sum_{j=1}^{n_i} \psi(\frac{X_{ij} - \tilde{\nu}}{\hat{\sigma}_n})$, $\bar{\psi}_{\cdot\cdot}(\boldsymbol{X}) \equiv \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \psi(\frac{X_{ij} - \tilde{\nu}}{\hat{\sigma}_n})$ とおき

$$T_M \equiv \sum_{i=1}^{k} n_i \{\bar{\psi}_{i\cdot}(\boldsymbol{X}) - \bar{\psi}_{\cdot\cdot}(\boldsymbol{X})\}^2$$

が検定統計量である．H_0 の下で $X_{11}, \cdots, X_{kn_k}$ は同一の連続型分布に従い, (3.8) より, H_0 の下で順序統計量 $\boldsymbol{X}_{(\cdot)} \equiv (X_{(1)}, \cdots, X_{(n)}) = \boldsymbol{x}_{(\cdot)} \equiv (x_{(1)}, \cdots, x_{(n)})$ を与えたときの T_M の条件付き分布は

$$P_0(T_M \leq t|\boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}) = \frac{1}{n!}\#\{\boldsymbol{v} : \sum_{i=1}^{k} n_i\{\bar{\psi}_{i\cdot}(\boldsymbol{v}) - \bar{\psi}_{\cdot\cdot}(\boldsymbol{v})\}^2 \leq t,\ \boldsymbol{v} \in \mathcal{V}_n\}$$

によって有意確率を計算する, ただし, $\boldsymbol{v} = (v_{11}, \cdots, v_{kn_k})$, $\mathcal{V}_n$ は (3.7) で定義したもの. また,

$$Z_M \equiv \frac{n-1}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{\psi(\frac{X_{ij}-\tilde{\nu}}{\hat{\sigma}_n}) - \bar{\psi}_{\cdot\cdot}(\boldsymbol{X})\}^2} \cdot T_M$$

とおけば, 定理 10.2 に対応する定理が導け (演習 10.9), 定理 10.3 と同様に $n \to \infty$ として, H_0 の下で $Z_M \xrightarrow{\mathcal{L}} Z \sim \chi^2_{k-1}$, すなわち, Z_M は自由度 $k-1$ のカイ二乗分布に分布収束する (詳細な証明は Shiraishi (1996)). $T(\boldsymbol{X}) = T_M$ が大きいとき H_0 を棄却する. $0 < \alpha < 1$ に対して $P_0(T_M > t_\alpha|\boldsymbol{X}^{(\cdot)} = \boldsymbol{x}^{(\cdot)}) \leq \alpha$ かつ $P_0(T_M \geq t_\alpha|\boldsymbol{X}^{(\cdot)} = \boldsymbol{x}^{(\cdot)}) > \alpha$ となる t_α を探す. このとき, 有意水準 α の検定方式は検定関数 $\phi(\cdot)$ を使って,

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & (T_M > t_\alpha\text{のとき}) \\ \dfrac{\alpha - P_0(T_M > t_\alpha|\boldsymbol{X}^{(\cdot)} = \boldsymbol{x}^{(\cdot)})}{P_0(T_M = t_\alpha|\boldsymbol{X}^{(\cdot)} = \boldsymbol{x}^{(\cdot)})} & (T_M = t_\alpha\text{のとき}) \\ 0 & (T_M < t_\alpha\text{のとき}) \end{cases}$$

と表現される. $n_1, \cdots, n_k$ が大きいとき, 検定方式は単純で検定関数 $\phi(\cdot)$ を使って,

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & (Z_M > \chi^2_{k-1,\alpha}\text{のとき}) \\ 0 & (Z_M < \chi^2_{k-1,\alpha}\text{のとき}) \end{cases}$$

と表現される.

[8] **M 推定量**; $i \neq i'$ に対して $\tilde{\nu}_{ii'} \equiv \dfrac{n_i\bar{X}_{i\cdot} + n_{i'}\bar{X}_{i'\cdot}}{n_i + n_{i'}}$,

$\hat{\sigma}_n \equiv \dfrac{\sqrt{\pi}}{\sqrt{2}\cdot n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}|X_{ij} - \bar{X}_{i\cdot}|$ とし

$$T_{ii'}(\theta) \equiv \frac{1}{n_i}\sum_{j=1}^{n_i}\psi(\frac{X_{ij}-\tilde{\nu}_{ii'}-\theta}{\hat{\sigma}_n}) - \frac{1}{n_{i'}}\sum_{j=1}^{n_{i'}}\psi(\frac{X_{i'j}-\tilde{\nu}_{ii'}+(\frac{n_i}{n_{i'}})\cdot\theta}{\hat{\sigma}_n})$$

とおく. $T_{ii'}(\theta)=0$ の解を $\breve{\theta}_{ii'}$ とすれば, 6.3 節 [7] より $\breve{\delta}_{ii'}=(1+\dfrac{n_i}{n_{i'}})\cdot\breve{\theta}_{ii'}$ は $\mu_i-\mu_{i'}$ の M 推定量である. また, $\alpha_i=\dfrac{1}{n}\sum_{i'=1}^{k} n_{i'}(\mu_i-\mu_{i'})$ より, α_i の M 推定量として

$$\breve{\alpha}_i=\frac{1}{n}\sum_{i'=1}^{k} n_{i'}\breve{\delta}_{ii'} \quad (i=1,\cdots,k)$$

が提案できる, ただし $\breve{\delta}_{ii}=0$ とする. i,j に対して $Y_{ij}\equiv X_{ij}-\breve{\alpha}_i$ とおき, M 統計量を $T_M(\theta)=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\psi(\dfrac{Y_{ij}-\theta}{\hat{\sigma}_n})$ とする. $T_M(\theta)=0$ の θ の解 $\breve{\nu}=\hat{\theta}$ が ν の M 推定量である. μ_i の M 推定量は

$$\breve{\mu}_i=\breve{\nu}+\breve{\alpha}_i \quad (i=1,\cdots,k).$$

[6] と同様に, 推定量の理論的正当性として, $\breve{\alpha}_i$ は密度関数 $f(x)$ に対称性を必要としないが, $\breve{\nu}$, $\breve{\mu}_i$ には $f(x)$ の対称性を必要とする.

密度関数 $f(x)$ に対称性を仮定できない場合は, $\tilde{\nu}=\bar{X}_{\cdot\cdot}$ を ν の推定量とする. このとき, μ_i の推定量は

$$\breve{\mu}_i=\tilde{\nu}+\breve{\alpha}_i \quad (i=1,\cdots,k).$$

10.6　データ解析

第一標本は現在もタバコを吸い続けている成人男子 6 人の 1 秒間における肺活量 (単位リットル), 第二標本は二年以内にタバコを吸うことをやめた成人男子 6 人の肺活量, 第三標本は二年以上前にタバコを吸うことをやめた成人男子 6 人の肺活量, 第四標本は今までにタバコを吸ったことのない非喫煙成人男子 6 人の肺活量を測定したものである. 観測結果は

タバコを吸い続けている	2.98	2.95	2.15	3.41	3.97	3.86
二年以内にタバコをやめた	3.54	4.40	3.28	2.28	3.34	3.92
二年以上前にタバコをやめた	3.69	3.90	3.82	4.08	3.76	4.38
タバコを吸ったことのない	4.41	4.96	3.50	3.66	4.68	4.11

である. 10.3, 10.4, 10.5 節の正規母集団での最良手法, ノンパラメトリック法, セミパラメトリック法を使って解析する. 解析プログラムとして『統計プログラミング』のソフトを使用した.

(i) 正規母集団での最良手法による結果

級内変動 6.788 総変動 10.477
級間変動 3.689
F 検定統計量の値 3.623
F 検定の p 値 0.031
一様性の帰無仮説は有意水準 5 パーセントで棄却された
第一標本の主効果 -0.49 平均 3.22
第二標本の主効果 -0.25 平均 3.46
第三標本の主効果 0.23 平均 3.94
第四標本の主効果 0.51 平均 4.22
全平均 3.71 共通分散 0.34

(ii) ノンパラメトリック法による結果

順位検定統計量の値 375.67 正規化順位検定統計量の値 7.513
順位検定の p 値 0.053
一様性の帰無仮説は有意水準 5 パーセントで棄却されなかった
第一標本の主効果 -0.51 平均 3.20
第二標本の主効果 -0.22 平均 3.49
第三標本の主効果 0.21 平均 3.93
第四標本の主効果 0.52 平均 4.23
全平均 3.71

(iii) セミパラメトリック法による結果

M 検定統計量の値	6.106	正規化 M 検定統計量の値	8.463
M 検定の p 値	0.037		

一様性の帰無仮説は有意水準 5 パーセントで棄却された

第一標本の主効果	-0.45	平均	3.29
第二標本の主効果	-0.23	平均	3.51
第三標本の主効果	0.20	平均	3.94
第四標本の主効果	0.47	平均	4.22
全平均	3.74		

10.7　傾向性と分散の異なるモデルの解析

10.1 節のモデルで位置パラメータに傾向性の制約

$$\mu_1 \le \mu_2 \le \cdots \le \mu_k \quad \Longleftrightarrow \quad \alpha_1 \le \alpha_2 \le \cdots \le \alpha_k \tag{10.5}$$

がある場合での統計解析法が論じられる．興味ある帰無仮説と対立仮説は,

$$\begin{cases} \text{帰無仮説} \quad H_0:\ \mu_1 = \cdots = \mu_k \\ \text{対立仮説} \quad H_1^*:\ \mu_1 \le \mu_2 \le \cdots \le \mu_k \ (\text{少なくとも 1 つの不等号は } <) \end{cases}$$

である．また, 傾向性の制約が正しいかどうかを確かめるためにもう 1 つの帰無仮説と対立仮説は,

$$\begin{cases} \text{帰無仮説} \quad H_0^{\#}:\ \mu_1 \le \mu_2 \le \cdots \le \mu_k \\ \text{対立仮説} \quad H_1^{\#}:\ \text{ある } i \text{ について } \mu_i > \mu_{i+1} \end{cases}$$

である．さらに, 傾向性の制約 (10.5) の下でのパラメータ $(\alpha_1, \cdots, \alpha_k)$, ν, $(\mu_1, \cdots, \mu_k)$ の点推定に興味がある．このとき, ロバートソン等 (Robertson et al (1988)) で正規分布による手法と順位検定法がレビューされている．

解析チャート

[1] 正規性の検定, [2] 分布の探索, 経験分布関数と正規分布関数の重ねかきグラフ

↓

① [1] で正規性が棄却されず, [2] で正規分布が選択され, 正規性が妥当と認められれば [3],[4] を選択 ② [2] で両側指数分布またはワイブル分布, 対数正規分布, 指数分布が選択されれば [5],[6] を選択 ③ これら以外であれば [7],[8] を選択

演習問題 10

10.1. 10.3 節の設定で, 対立仮説 H_1 の下で, T_S は自由度 $(k-1, n-k)$ 非心度 $\delta^2 \equiv \dfrac{1}{\sigma^2}\sum_{i=1}^{k} n_i(\mu_i - \bar{\mu}.)^2$ の非心 F 分布に従うことを示せ, ただし $\bar{\mu}. \equiv \sum_{i=1}^{k} n_i\mu_i/n$.

10.2. $X_{11}, \cdots, X_{kn_k}$ は互いに独立で, 各 X_{ij} は同一の分布関数 $F(\dfrac{x-\mu}{\sigma})$ (一般性を失うことなく $\int_{-\infty}^{\infty} x dF(x) = 0$, $\int_{-\infty}^{\infty} x^2 dF(x) = 1$ とする) をもち, $\lim_{n\to\infty} \dfrac{n_i}{n} = \lambda_i > 0$ $(i = 1, \cdots, k)$ ならば, $n \to \infty$ として

$$(k-1)T_S \xrightarrow{\mathcal{L}} \sum_{i=1}^{k} \lambda_i (Y_i - \sum_{j=1}^{k} \lambda_j Y_j)^2 \ \sim\ \chi^2_{k-1}$$

が成り立つことを示せ, ただし, $Y_1, \cdots, Y_k$ は互いに独立で $Y_i \sim N(0, \dfrac{1}{\lambda_i})$.

10.3. $X_{11}, \cdots, X_{kn_k}$ は互いに独立で, 各 X_{ij} は, 分布関数 $F(\dfrac{x-\mu_i}{\sigma})$ をもつとする. さらに $\int_{-\infty}^{\infty} x dF(x) = 0$ を仮定するならば, $E(\tilde{\alpha}_i) = \alpha_i$, $E(\tilde{\nu}) = \nu$, $E(\tilde{\mu}_i) = \mu_i$ を示せ.

10.4. $X_{11}, \cdots, X_{kn_k}$ は互いに独立で, 各 X_{ij} は, 分布関数 $F(\frac{x-\mu_i}{\sigma})$ をもつとする. さらに $\int_{-\infty}^{\infty} x dF(x) = 0$, $\int_{-\infty}^{\infty} x^2 dF(x) = 1$ を仮定し, $\lim_{n\to\infty} \frac{n_i}{n} = \lambda_i > 0$ $(i = 1, \cdots, k)$ ならば, $n \to \infty$ として $\tilde{\alpha}_i \xrightarrow{P} \alpha_i$, $\tilde{\nu} \xrightarrow{P} \nu$, $\tilde{\mu}_i \xrightarrow{P} \mu_i$ を示せ.

10.5. H_0 の下で $P_0(T_R = t) > 0$ ならば, $P_0(T_R = t) \geq \frac{n_1! \cdots n_k!}{n!}$ であることを示せ.

10.6. 帰無仮説 H_0 の下で [3] で定義した $\boldsymbol{S}$ の平均と分散共分散行列はそれぞれ

$$E_0(\boldsymbol{S}) = \boldsymbol{0}, \quad V_0(\boldsymbol{S}) = \mathrm{diag}(\frac{n}{n_1}, \cdots, \frac{n}{n_k}) - \boldsymbol{1}_k \cdot {}^t\boldsymbol{1}_k$$

となることを示せ.

10.7. 10.5 節の設定で M 検定統計量 T_M は H_0 の下で $P_0(T_M = t | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}) > 0$ ならば, $P_0(T_M = t | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}) \geq \frac{n_1! \cdots n_k!}{n!}$ であることを示せ.

10.8. 10.5 節の設定で, $S_i' \equiv \sqrt{\frac{n(n-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{\psi(\frac{X_{ij}-\tilde{\nu}}{\hat{\sigma}_n}) - \bar{\psi}_{\cdot\cdot}(\boldsymbol{X})\}^2}} \cdot \{\bar{\psi}_{i\cdot}(\boldsymbol{X}) - \bar{\psi}_{\cdot\cdot}(\boldsymbol{X})\}$ $(i = 1, \cdots, k)$ とおく. H_0 の下で順序統計量 $\boldsymbol{X}_{(\cdot)} \equiv (X_{(1)}, \cdots, X_{(n)}) = \boldsymbol{x}_{(\cdot)} \equiv (x_{(1)}, \cdots, x_{(n)})$ を与えたときの H_0 の下で $\boldsymbol{S}' \equiv {}^t(S_1', \cdots, S_k')$ の条件付き平均と分散共分散行列はそれぞれ

$$E_0(\boldsymbol{S}' | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}) = {}^t(E_0\{S_1' | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}\}, \cdots, E_0\{S_k' | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}\}) = \boldsymbol{0},$$

$$\begin{aligned} V_0(\boldsymbol{S}' | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}) &\equiv (\mathrm{Cov}_0\{S_i', S_j' | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}\})_{i,j=1,\cdots,k} \\ &= \mathrm{diag}(\frac{n}{n_1}, \cdots, \frac{n}{n_k}) - \boldsymbol{1}_k \cdot {}^t\boldsymbol{1}_k, \end{aligned}$$

となることを示せ, ただし

$$\mathrm{Cov}_0\{S_i', S_j' | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}\} = E_0\{S_i' S_j' | \boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}\}$$

$$-E_0\{S_i'|\boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}\}E_0\{S_j'|\boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)}\}.$$

10.9. 10.8 の記号で $T_M = {}^t\boldsymbol{S}'\{V_0(\boldsymbol{S}'|\boldsymbol{X}_{(\cdot)} = \boldsymbol{x}_{(\cdot)})\}^-\boldsymbol{S}'$ が成り立つことを示せ.

10.10. 10.5 節の $\psi(x)$ で $b = \infty$ すなわち, $\psi(x) \equiv x$ とすれば, Z_M は T_S の狭義増加関数であることを示せ.

10.11. $\psi(x) \equiv x$ とすれば, T_M に基づく検定は並べ替え F 検定 (並べ替え t 検定と同じ考え方の検定) と同値であることを示せ.

10.12. $\psi(x) \equiv x$ とすれば, 10.5 節 [8] の M 推定量は $\breve{\alpha}_i = \tilde{\alpha}_i$, $\breve{\nu} = \tilde{\nu}$, $\breve{\mu}_i = \tilde{\mu}_i$ を示せ.

10.13. (ウェルチの検定統計量の拡張) $X_{11}, \cdots, X_{kn_k}$ は互いに独立で, 各 X_{ij} は分布関数 $F(\dfrac{x-\mu}{\sigma_i})$ (一般性を失うことなく $\displaystyle\int_{-\infty}^{\infty} xdF(x) = 0$, $\displaystyle\int_{-\infty}^{\infty} x^2dF(x) = 1$ とする) をもつとする. $\tilde{\sigma}_i^2 \equiv \dfrac{1}{n_i - 1}\displaystyle\sum_{j=1}^{n_1}(X_{ij} - \bar{X}_{i\cdot})^2$,

$\bar{\bar{X}}_{\cdot\cdot} \equiv \displaystyle\sum_{i=1}^{k}\frac{n_i \cdot \bar{X}_{i\cdot}}{\tilde{\sigma}_i^2} \Big/ \sum_{i=1}^{k}\frac{n_i}{\tilde{\sigma}_i^2}$ で定義し $T' \equiv \dfrac{n_i(\bar{X}_{i\cdot} - \bar{\bar{X}}_{\cdot\cdot})^2}{\tilde{\sigma}_i^2}$ とおく. このとき, $\displaystyle\lim_{n\to\infty}\frac{n_i}{n} = \lambda_i > 0$ $(i = 1, \cdots, k)$ ならば, $n \to \infty$ として T' は自由度 $k-1$ のカイ二乗分布に分布収束することを示せ.