人文 (23) 略解

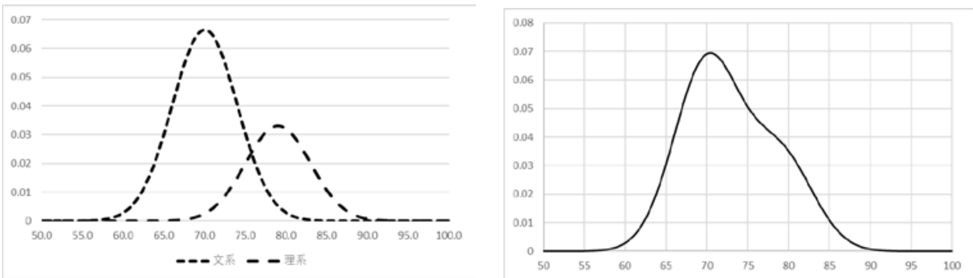
人文１

[1]　 A さんの点数の文系受験者の中での偏差値は 60，B さんの点数の理系受験者の中での偏差値は 67.5 である。受験生全体での合格率は約 0.66 となる。

[2]　 混合分布の期待値と分散は $\xi = 73.0$ および $\tau^2 = 34.0$ である，受験生全体では， A さんの偏差値は約 51.7，B さんの偏差値は約 72.3 となる。

[3]

[3-1] 混合比を考慮した文系のみと理系のみ，および受験生全体の点数の分布は図 1 のようである。



(a) 文系のみと理系のみ　　(b) 受験生全体

図 1：試験の点数の分布

[3-2] 受験生全体の点数の分布の確率密度関数は

$$g(x) = \frac{1}{3\sqrt{32\pi}}\left\{2\exp\left(-\frac{(x-70)^2}{32}\right) + \exp\left(-\frac{(x-79)^2}{32}\right)\right\}$$

である。これを x で微分すると，

$$g'(x) = \frac{1}{3\sqrt{32\pi}}\left\{2\times\frac{-2(x-70)}{32}\exp\left(-\frac{(x-70)^2}{32}\right) + \frac{-2(x-79)}{32}\exp\left(-\frac{(x-79)^2}{32}\right)\right\} \qquad (1)$$

となり，$x < 70$ のとき $g'(x) > 0$ および $x > 79$ のとき $g'(x) < 0$ であるので，$g'(x) = 0$ となる最頻値 x_{mode} は $70 < x_{mode} < 79$ の範囲内である。

人文２

[1]　偏相関係数の値は $r_{XY/W}=0$ である。偏相関係数の値は 0 であるので，偏差値 W の影響を除くと X と Y とは無相関であった。

[2]

[2-1] $V[E_1]=1-r_{XW}^2$， $V[E_2]=1-r_{YW}^2$。

[2-2] E_1 と E_2 の共分散は $Cov[E_1,E_2]=r_{XY}-r_{YW}r_{XW}$ であるので，E_1 と E_2 の間の相関係数は $R[E_1,E_2]=\dfrac{r_{XY}-r_{YW}r_{XW}}{\sqrt{(1-r_{XW}{}^2)(1-r_{YW}{}^2)}}$ と偏相関係数 $r_{XY/W}$ に一致する。

[3]　$a = Cov[X, Y]/V[X] = r_{XY} = 0.72$ であり $\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}=\begin{pmatrix} 0 \\ 0.8 \end{pmatrix}$ となる。

[4]

[4-1] X_A と Y_A の間の相関係数は 0 であり，T_A と X_A の間の相関係数は $\sqrt{2}/2$ となる。T_A と Y_A の間の相関係数も同様に $\sqrt{2}/2$ である。

[4-2] 合格者のみでの X_A と Y_A の間には負の相関が観測される。

人文３

[1]　マハラノビスの平方距離はそれぞれ $D_1^2 = 2$，$D_2^2 = 1/2$ となる。

[2]　$f(x) = 0.75\boldsymbol{x}^T\boldsymbol{x} - 2.5\mathbf{1}^T\boldsymbol{x} + 1.5$。

[3]

[3-1] G_1 では $Y_1 \sim N(\sqrt{2}, 1)$，G_2 では $Y_1 \sim N(-\sqrt{2}, 2^2)$ となる。

[3-2] $P(Y_1 < 0 \mid G_1) \approx P(Z < -1.41) \approx 0.0793$ および $P(Y_1 \geq 0 \mid G_2) \approx P(Z \geq 0.71) \approx 0.2389$ となる。

[4]　判別境界を構成した観測値を再度用いて誤判別率を求めた場合，誤判別率は実際より低く評価される。そのため，構成の際に用いなかった観測値を利用して誤判別率を計算または推測することが好ましい。

人文４（社会４と同じ）

[1]　表 1 より求めた回帰式は，男性では，$y = 4.2 + 0.6x$，女性では $y = 3.6 + 0.6x$ である。両回帰式は平行であり，同じ労働時間であっても男性のほうが平均 0.6 千円賃金が高い。

[2]　$\bar{y}_M / \bar{x}_M = 1.2$，$\bar{y}_F / \bar{x}_F = 1.2$ と 1 時間当たりの賃金は男女で同じになる。よって，この指標でみると男女の差は見られない。

[3]　回帰係数は，男性では $d_M \approx 0.4167$，$c_M = 3.5$ となり，女性では $d_F \approx 0.4167$，$c_F = 3.0$ となる。同じ 7.2 千円を得ている人では，男性では平均 6.5 時間働き，女性では平均 6.0 時間働いている。

[4]　上問 [1]~[3] の結果から，統計家 A は女性にとっての不公平さを認め，逆に統計家 C は男性にとっての不公平さを認めている。その一方で統計家 B は男女間で不公平はないという結論となった。

社会 (23) 略解

社会 1

[1]　平均：$\bar{x} \approx 3.73$，$\bar{y} \approx 27.49$，分散：${s_x}^2 \approx 2.33$，${s_y}^2 \approx 102.63$，共分散：$s_{xy} \approx 6.69$，相関係数：$r \approx 0.43$，比：$\hat{R} \approx 7.37$。

[2]　$g(x, y) = y/x$ とすると，$\dfrac{\partial}{\partial x}\dfrac{y}{x} = -\dfrac{y}{x^2}$，$\dfrac{\partial}{\partial y}\dfrac{y}{x} = \dfrac{1}{x}$ より $V[\hat{R}] \approx \dfrac{1}{n{\mu_X}^2} E[(Y - RX)^2]$ となる。

[3]　$\hat{\mu}_Y = \hat{R}\mu_X$ とすると

$$
\begin{aligned}
V[\hat{\mu}_Y] &= \frac{{\mu_Y}^2}{n}\left\{\left(\frac{\sigma_X}{\mu_X}\right)^2 + \left(\frac{\sigma_Y}{\mu_Y}\right)^2 - 2\rho\frac{\sigma_X}{\mu_X}\frac{\sigma_Y}{\mu_Y}\right\} \\
&= \frac{{\mu_Y}^2}{n}\{\mathrm{cv}(X)^2 + \mathrm{cv}(Y)^2 - 2\rho \cdot \mathrm{cv}(X) \cdot \mathrm{cv}(Y)\}
\end{aligned}
$$

となる。

[4]　比推定量 $\hat{R} = \bar{y} / \bar{x}$ のほうが $\bar{y} / \mu_X$ よりも分散が小さい条件は，

$$
\rho > \frac{R\sigma_X}{2\sigma_Y} = \frac{\sigma_X / \mu_X}{2\sigma_Y / \mu_Y} = \frac{\mathrm{cv}(X)}{2\mathrm{cv}(Y)}
$$

となる。よって正解は (b) である。また，μ_X が既知であれば，$\hat{R}$ よりも $\bar{y} / \mu_X$ のほうが R の推定量としては推奨される。

社会２

[1]　X の期待値は $E[X]=\dfrac{ab}{a-1}$ $(a>1)$ と求められる。累積分布関数は，$x>b$ のとき $F(x)=1-\left(\dfrac{b}{x}\right)^{a}$ である。よって中央値 m は $m=2^{1/a}b$ となる。

[2]　$x>c$ のときの条件付き確率密度関数は $f(x\mid x>c)=\dfrac{1}{P(X>c)}\dfrac{ab^{a}}{x^{a+1}}=\dfrac{ac^{a}}{x^{a+1}}$ となる。確率密度関数 $f(x)$ は，$x>b$ の条件付き分布とも考えられることから，条件 c $(c>b)$ がいくらであっても関数形は同じである。

[3]　$a=\log 5/\log 4\approx 1.16$。

[4]　A さんは 4 百万円の契約を取る必要がある。また，上位 4%となる x の値は 16 百万円である。

社会３

[1]　積分計算により

$$E[\exp(kU)] = \frac{1}{\sqrt{2\pi\sigma_u^2}}\int_{-\infty}^{\infty}\exp(ku)\exp\left[-\frac{u^2}{2\sigma_u^2}\right]du = \exp\left(\frac{k^2}{2}\sigma_u^2\right)$$

と求められる。

[2]　マルチンゲールとなる条件は $\alpha_0 = -\sigma_u^2/2$ および $\alpha_1 = 1$ である。

[3]　予測分散は，$E[(Y_{T+1} - \tilde{Y}_{T+1})^2 \mid Y_T] = Y_T^2\{\exp(\sigma_u^2) - 1\}$ と求められる。

[4]　α_1 の最小二乗 (OLS) 推定量を $\hat{\alpha}_1$ とするとき，$|\alpha_1| < 1$ の場合，$\sqrt{T}(\hat{\alpha}_1 - \alpha_1)$ は $N(0,(1-\alpha_1^2))$ に分布収束する。しかし，$\alpha_1 = 1$ の場合はこの漸近分布が退化し，$\sqrt{T}(\hat{\alpha}_1 - \alpha_1)$ は 0 に確率収束するので，OLS 推定量に対する通常の t 検定は適切ではない。

社会４（人文４と同じ）

理工 (23) 略解

理工１

[1]

[1-1]

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} (y_1 + y_2 + y_3 + y_4)/4 \\ (y_1 + y_2 - y_3 - y_4)/4 \\ (y_1 - y_2 + y_3 - y_4)/(4s) \end{pmatrix}$$

であり，

$$V[\hat{\boldsymbol{\beta}}] = \begin{pmatrix} \sigma^2/4 & 0 & 0 \\ 0 & \sigma^2/4 & 0 \\ 0 & 0 & \sigma^2/(4s^2) \end{pmatrix}$$

となる。

[1-2] $s = L$

[1-3] $s = L$

[2]

[2-1] $t = 5$

[2-2] $t = 3$

理工２

[1]　部分積分により $E[X]=\int_0^\infty \exp\left(-\frac{x^2}{2\sigma^2}\right)dx$ となり，$\frac{2}{\sqrt{2\pi}\sigma}\int_0^\infty \exp\left(-\frac{x^2}{2\sigma^2}\right)dx=1$ である の で，$E[X]=\sqrt{\frac{\pi}{2}}\sigma\approx 1.25\sigma$ となる。中央値は，$\sqrt{2\log 2}\sigma\approx 1.18\sigma$ となり, 最頻値は $x=\sigma$ となる。

[2]　U と V の同時確率密度関数 $g(u,v)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{u^2+v^2}{2\sigma^2}\right)$ で極座標変換

$$\begin{cases} u = x\cos\theta \\ v = x\sin\theta \end{cases}$$

を行うと，$g(u,v)dudv=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)xdxd\theta$ となり，x の周辺分布を求めると $f(x)=\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)$ と，パラメータ σ のレイリー分布の確率密度関数が得られる。

[3]　最尤推定量は $\hat{\sigma}=\sqrt{\frac{1}{2n}\sum_{i=1}^n X_i{}^2}$ となる。

[4]　σ の最尤推定値は $\hat{\sigma}=\sqrt{\frac{W}{2n}}=\sqrt{\frac{89.96}{20}}\approx 2.12$ となる。また，σ の 95%信頼区間は

$$\left(\sqrt{\frac{89.96}{34.17}},\sqrt{\frac{89.96}{9.59}}\right)\approx(1.62, 3.06)$$

となる。

理工３

[1]　期待値は $\xi = p\mu_1 + (1-p)\mu_2$ である。分散は $\tau^2 = \sigma^2 + p(1-p)(\mu_1 - \mu_2)^2$ であり，それが最大となるのは $p = 0.5$ の場合である。

[2]　$\gamma(X) = pf_1(X)/g(X)$ であるので，

$$E[\gamma(X)] = \int_{-\infty}^{\infty} \gamma(x)g(x)dx = p\int_{-\infty}^{\infty} f_1(x)dx = p\,,$$

$$\frac{1}{p}E[\gamma(X)X] = \int_{-\infty}^{\infty} xf_1(x)dx = \mu_1\,,$$

$$\frac{1}{1-p}E[\{1-\gamma(X)\}X] = \int_{-\infty}^{\infty} xf_2(x)dx = \mu_2$$

となる。

[3]　$$\gamma(x) = \frac{1}{1 + \{(1-p)/p\}\exp[-(\mu_1 - \mu_2)\{x - (\mu_1 + \mu_2)/2\}/\sigma^2]}$$

[4]　初期値 $\hat{p}^{(0)}$，$\hat{\mu}_1^{(0)}$，$\hat{\mu}_2^{(0)}$ から始まり，t 回目の反復における $\gamma(x_i)$ の値を

$$\gamma^{(t)}(x_i) = \frac{1}{1 + \{(1-\hat{p}^{(t-1)})/\hat{p}^{(t-1)}\}\exp[-(\hat{\mu}_1^{(t-1)} - \hat{\mu}_2^{(t-1)})\{x_i - (\hat{\mu}_1^{(t-1)} + \hat{\mu}_2^{(t-1)})/2\}/\sigma^2]}$$

とし，各パラメータの反復値を

$$\hat{p}^{(t)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i^{(t)}$$

$$\hat{\mu}_1^{(t)} = \frac{1}{\hat{p}^{(t)}}\cdot\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i^{(t)}x_i,\quad \hat{\mu}_2^{(t)} = \frac{1}{1-\hat{p}^{(t)}}\cdot\frac{1}{n}\sum_{i=1}^{n}(1-\hat{\gamma}_i^{(t)})x_i$$

とする。この反復を収束するまで繰り返す。

理工４

[1]　$a(y) = y,\ b(\lambda) = \log \lambda,\ c(\lambda) = -\lambda,\ d(y) = -\log y!$

である。期待値は

$$E[a(Y)] = E[Y] = -c'(\lambda)/b'(\lambda) = -(-\lambda)'/(\log \lambda)' = 1/(1/\lambda) = \lambda$$

であり，分散は

$$V[a(Y)] = V[Y] = \{b''(\lambda)c'(\lambda) - c''(\lambda)b'(\lambda)\}/\{b'(\lambda)\}^3$$
$$= \frac{(-1/\lambda^2)(-1) - 0}{(1/\lambda)^3} = \lambda$$

と求められる。

[2]　$\lambda \approx 2.53$

[3]　尤度方程式は $l'(\beta) = \sum_{i=1}^{n} x_i\{y_i - \exp(\beta x_i)\} = 0$ となる。

[4]　ニュートン法の反復スキームは

$$\beta^{(t+1)} = \beta^{(t)} - \frac{l'(\beta^{(t)})}{l''(\beta^{(t)})} = \beta^{(t)} + \frac{\sum_{i=1}^{n} x_i\{y_i - \exp(\beta^{(t)} x_i)\}}{\sum_{i=1}^{n} \{x_i^2 \exp(\beta^{(t)} x_i)\}}$$

となる。

医薬 (23) 略解

医薬 1

[1]　T 群の対数オッズは $\alpha+\beta$ であり，C 群の対数オッズは α となる。また，T 群の C 群に対するオッズ比は $\exp(\beta)$ となる。

[2]　$c_1 = 11$，$c_2 = 7$，$c_3 = -10$，$c_4 = -10$

[3]　(1) $\hat{\alpha} \approx -0.4055$，(2) $\hat{\beta} \approx 1.253$

[4]　尤度比カイ二乗検定統計量の値は $\chi^2 = 2.1$ である。自由度 1 のカイ二乗分布の上側 5%点は 3.84 であるので，$H_0 : \gamma = 0$ は棄却されない。

[5]　陽性的中率は $\dfrac{2+3+1}{2+5+2} = \dfrac{6}{9} = \dfrac{2}{3} \approx 0.667$，陰性的中率は $\dfrac{1+0+5}{3+2+6} = \dfrac{6}{11} \approx 0.545$ となる。

医薬２

[1]　最尤推定値は $\hat{\lambda} = 1/\bar{t}$ となる。

[2]　フィッシャー情報量は，$I_n(\lambda) = \dfrac{n}{\lambda^2}$ となる。

[3]　$\mu = z_\alpha + z_\beta$

[4]　$n = \dfrac{\lambda_1{}^2 (z_\alpha + z_\beta)^2}{(\lambda_1 - \lambda_0)^2}$

[5]　$n = 32$

医薬３

[1] $l(\theta) = -\frac{1}{2}\left\{\sum_{k=1}^{K}\log(2\pi{\sigma_k}^2) + \sum_{k=1}^{K} w_k (y_k - \theta)^2\right\}$

[2] 最尤推定量は $\hat{\theta} = \frac{\sum_{k=1}^{K}(1/{\sigma_k}^2)Y_k}{\sum_{k=1}^{K}(1/{\sigma_k}^2)}$ であり，期待値と分散はそれぞれ $E[\hat{\theta}] = \theta$，

$V[\hat{\theta}] = \frac{1}{\sum_{k=1}^{K} w_k}$ となる。

[3] $E[Q] = \sum_{k=1}^{K} w_k V[Y_k] - \frac{\sum_{k=1}^{K} {w_k}^2 V[Y_k]}{\sum_{k=1}^{K} w_k}$

[4] $\hat{\tau}^2 = \frac{Q - (K-1)}{\sum_{k=1}^{K} w_k - \left(\sum_{k=1}^{K} {w_k}^2\right)\Big/\left(\sum_{k=1}^{K} w_k\right)}$

[5] $\frac{\tau^2}{\tau^2 + \sigma^2} = \frac{7.15 - 4 + 1}{7.15} \approx 0.58$

医薬４

[1]　$\hat{\boldsymbol{\mu}} = (\hat{\mu}_0, \hat{\mu}_4, \hat{\mu}_8)^T$ は 3 変量正規分布に従い，その平均ベクトルと分散共分散行列はそれぞれ $E[\hat{\boldsymbol{\mu}}] = \boldsymbol{\mu}$，$V[\hat{\boldsymbol{\mu}}] = \frac{1}{N}\Sigma$ となる。

[2]　Toeplitz では AIC は 424.84，無構造では AIC は 427.74 となる。AIC 最小という規準からは Toeplitz のほうが適切と判断されるが，差は 2.9 とあまり大きくないので，必ずしも Toeplitz のほうがよいとまではいえないという判断もあり得る。

[3]　$\hat{R} = \frac{1}{2.56}\begin{pmatrix} 2.56 & 1.54 & 1.48 \\ 1.54 & 2.56 & 1.54 \\ 1.48 & 1.54 & 2.56 \end{pmatrix} \approx \begin{pmatrix} 1 & 0.60 & 0.58 \\ 0.60 & 1 & 0.60 \\ 0.58 & 0.60 & 1 \end{pmatrix}$

[4]　変化量は $\hat{\mu}_8 - \hat{\mu}_0 = 5.72 - 5.18 = 0.54$ である。その分散の推定値は $\hat{V}[\hat{\mu}_8 - \hat{\mu}_0] \approx 0.054$ となり，よって，標準誤差は $\sqrt{0.054} \approx 0.23$ と求められる。

[5]　帰無仮説は $H_0 : \mu_8 - \mu_0 = 0$ と書けるので，検定統計量の値は $Z = 0.54/0.23 \approx 2.35$ となる。この値は標準正規分布の上側 2.5%の $z_{0.025} = 1.96$ よりも大きいので，有意水準 5%の両側検定で帰無仮説は棄却される。

共通 (23) 略解

モンティ・ホール問題としても有名な問題である。

[1]　ベイズの定理 $P(A_j \mid B) = \dfrac{P(B \mid A_j)P(A_j)}{P(B)}$ $(j = 1, . . . , m)$ において，$P(A_j) = 1/m$ であるので，$P(A_j \mid B) = \dfrac{1}{mP(B)} \times P(B \mid A_j) = c \cdot P(B \mid A_j)$ となる。

[2]　箱 j が当たりである事象を A_j とし，和夫さんが箱 1 を選んだとき，令美さんが箱 3 を開ける事象を B とすると，上問 [1] より

$$
\begin{aligned}
P(A_1 \mid B) : P(A_2 \mid B) : P(A_3 \mid B) &= P(B \mid A_1) : P(B \mid A_2) : P(B \mid A_3) \\
&= 1/2 : 1 : 0 = 1 : 2 : 0
\end{aligned}
$$

であるので，箱 2 が当たりの確率は 2/3 となる。

[3]　$P(B \mid A_1) = w$ とすると $p = P(A_2 \mid B) = \dfrac{1}{1+w}$ であり，p の存在範囲は $1/2 \le p \le 1$ となる。

[4]　試行回数が n のとき，$H_0 : p = 1/2$ の下で $P(X \ge n-1 \mid p = 1/2) = \dfrac{n+1}{2^n}$ であり，$\alpha =$ 0.05 で棄却される最小の n は $n_0 = 8$ で，検出力は $P(X \ge 7 \mid p = 2/3) \approx 0.195$ となる。別解として，二項分布の正規近似を用いた場合は $n_0 = 7$ となり，検出力は約 0.14 となる。