人文 (22) 略解

人文１

[1]　全体での得点 t の平均と分散は $\bar{t}=6.75$ および $s_t^{\,2}=5.4225$ となる。

[2]　和 w と差 d の平均および分散はそれぞれ $\bar{w}=13.5$， $\bar{d}=3.7$ および $s_w^{\,2}=4.64$， $s_d^{\,2}=3.36$ となる。

[3]　相関係数は $r_{wd}=0$ となる。

[4]　正規分布表から $P(U-V\geq 0)\approx 0.4562$ を得る。

人文２

[1]　G_1 では $Y\sim N(2,3)$ であり，G_2 では $Y\sim N(-2,3)$ となる。

[2]　$D_1^{\,2}=\dfrac{1}{0.75}(x_1^{\,2}+x_2^{\,2}-x_1x_2-x_1-x_2+1)$ および $D_2^{\,2}=\dfrac{1}{0.75}(x_1^{\,2}+x_2^{\,2}-x_1x_2+x_1+x_2+1)$ であり，$D_1^{\,2}=D_2^{\,2}$ と置いて $x_1+x_2=0$ を得る。

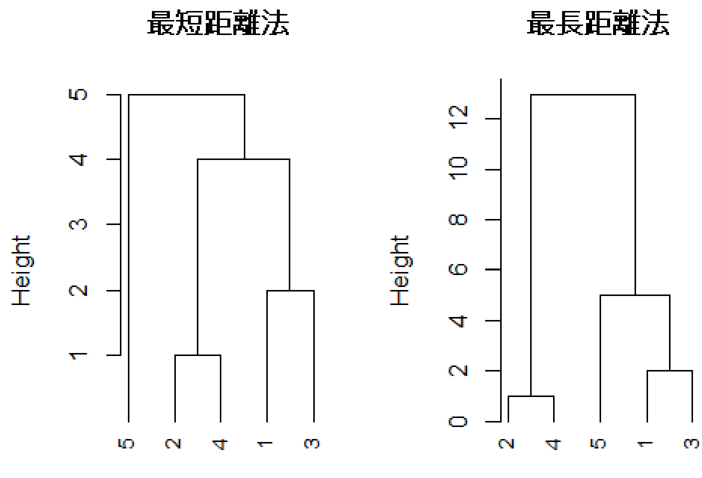
[3]　$Q=0.1251$。

[4]　$c=-\dfrac{3}{4}\log 2\approx -0.52$ で，$Q^*\approx 0.115$。

人文３

[1]

[1-1] それぞれのデンドログラムは次の図のようになる。



[1-2] 最短距離法のクラスターは {1, 2, 3, 4} と {5} に，最長距離法のクラスターは {1, 3, 5} と {2, 4} になる。

[1-3] 最短距離法のクラスターでは $W(C_1)+W(C_2)=35/4$ となり，最長距離法のクラスターでは $W(C_1)+W(C_2)=9/2$ となる。

[2]

[2-1] C_1 については $\bar{x}_{11}=1/2$, $\bar{x}_{12}=3/2$ となり，C_2 については $\bar{x}_{21}=2/3$, $\bar{x}_{22}=1$ となる。新たなクラスターは $C_1=\{1, 3, 5\}$, $C_2=\{2, 4\}$ となり，

$W(C_1)+W(C_2)=(2+5+5)/3+(1)/2=9/2$ である。

[2-2] 一般に，初期クラスターをランダムに決めることを何度か繰り返し，それぞれの中で最小になる結果を選択する。

人文４

[1] 測定の妥当性とは，測定すべきものがきちんと測定できているかを意味し，測定の信頼性とは，測定結果の安定性，一貫性を意味する。例えば，中学生の数学の能力を測るとき，テストで計算問題ばかりを出題したのでは，そのテストの信頼性は高いかも知れないが，それが生徒の数学の能力を真に測定しているとは限らない。

[2] $\alpha=\dfrac{m}{m-1}\left(1-\dfrac{\sum_{j=1}^{m}s_j^{\,2}}{s_T^{\,2}}\right)=\dfrac{m\bar{C}}{\bar{V}+(m-1)\bar{C}}$ であり，相関行列の場合は $\bar{V}=1$, $\bar{C}=\bar{R}$ であるので，$\alpha'=\dfrac{m\bar{R}}{1+(m-1)\bar{R}}$ を得る。$\alpha'=0$ となるのは $\bar{R}=0$ の場合で，$\alpha'=1$ となるのは項目間のすべての相関係数が 1 の場合に限る。

[3] $m\geq 4$ とすればよい。

[4]

[4-1] $\alpha'\approx 0.76$

[4-2] $\lambda\approx 0.66$

社会 (22) 略解

社会１

[1]　$E[\hat{T}_\pi]=E\left[\sum_{i=1}^{N} I_i(S)\frac{y_i}{\pi_i}\right]=\sum_{i=1}^{N}\pi_i\cdot\frac{y_i}{\pi_i}=\sum_{i=1}^{N} y_i=T_y$　より不偏性が示される。

[2]　$E[I_i(S)I_j(S)]=\pi_{ij}\quad(i,j=1,\ldots,N)$　および　$Cov[I_i(S),I_j(S)]=\pi_{ij}-\pi_i\pi_j$　より導かれる。

[3]　$E[\hat{V}_1[\hat{T}_\pi]]=\sum_{i=1}^{N}\sum_{j=1}^{N}(\pi_{ij}-\pi_i\pi_j)\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}=V[\hat{T}_\pi]$　より不偏性が示される。

[4]　$\sum_{i=1}^{N} I_i(S)=n$　および　$\sum_{i=1}^{N} I_i(S)I_j(S)=nI_j(S)$　より導かれる。

[5]　上問 [3] および [4] の結果から計算により導かれる。

社会２

[1]　$P_L\approx 97.18$

[2]

$$P_L=\frac{\sum_i p_{1i}q_{0i}}{\sum_i p_{0i}q_{0i}}\times 100,\quad P_P=\frac{\sum_i p_{1i}q_{1i}}{\sum_i p_{0i}q_{1i}}\times 100$$

$$Q_L=\frac{\sum_i p_{0i}q_{1i}}{\sum_i p_{0i}q_{0i}}\times 100,\quad Q_P=\frac{\sum_i p_{1i}q_{1i}}{\sum_i p_{1i}q_{0i}}\times 100$$

[3]　$\overline{x}=P_L$，　$\overline{y}=Q_L$

[4]　$s_{xy}=P_PQ_L-P_LQ_L=(P_P-P_L)Q_L$

[5]　一般に，価格上昇率の相対的に大きな商品の購入量を減らし，その小さな商品の購入量を増やすのが合理的な行動と考えられる。

社会３

[1]　$\tilde{\alpha}=\frac{X_TY_1-X_1Y_T}{X_T-X_1}$，　$\tilde{\beta}=\frac{Y_T-Y_1}{X_T-X_1}$　につき，計算により求める。

[2]　等号は，$X_1,\ldots,X_T$ 全体の平均が $m=(X_1+X_T)/2$ であり，$X_2=\cdots=X_{T-1}=m$ の場合である。

[3]　予測分散は，$V[\tilde{Y}_{T+1}]=\left\{\frac{\{X_{T+1}-(X_1+X_T)/2\}^2}{(X_T-X_1)^2/2}+\frac{3}{2}\right\}\sigma^2$　となる。

[4]　$X_c=(X_1+X_T)/2$　とすれば，1 次式モデルの成否が不明な場合には最良となる。

社会４

[1]　A : $a=0.8$，B : $a=-0.8$，C : $a=0$　である。

[2]　移動平均表現は　$y_t=v_t+a(v_{t-4}+ay_{t-8})=v_t+av_{t-4}+a^2v_{t-8}+\cdots$　となる。$E[y_t]=0$　であり，

$V[y_t]=\dfrac{\sigma^2}{1-a^2}$ となる。また，$\gamma(k)=E[y_t y_{t-k}]=\sigma^2\dfrac{a^h}{1-a^2}$ であり，自己相関関数は

$$\rho(k)=\begin{cases} a^h & (k=4h,\ h \text{ は自然数}) \\ 0 & (\text{それ以外}) \end{cases}$$

となる。

[3] $\hat{a}=\dfrac{\sum_{t=5}^{T} y_t y_{t-4}}{\sum_{t=5}^{T} y_{t-4}{}^2}$ であり，分散 σ^2 の推定量としては，$\hat{\sigma}^2=\dfrac{1}{T}\sum_{i=5}^{T}(y_t-\hat{a}y_{t-4})^2$ を用いる。

$\hat{\sigma}^2$ は σ^2 の一致推定量である。

[4] 予測量を $\hat{y}_{T+k|T}=\hat{a}^{s+1}y_{T+k-4(s+1)}$ によって与える。このとき予測誤差は

$$y_{T+k}-\hat{y}_{T+k|T}=(a^{s+1}-\hat{a}^{s+1})y_{T+k-4(s+1)}+\sum_{j=0}^{s}a^j v_{T+k-4j}$$

となる。

理工 (22) 略解

理工１

[1] $E[Z] = mp + nq$,　$V[Z] = mp(1-p) + nq(1-q)$

[2] $m_Z(\theta) = (pe^{\theta} + 1 - p)^m (qe^{\theta} + 1 - q)^n$

[3] $P(Z \geq 10) \leq \dfrac{2.455}{(7.5)^2} \approx 0.044$

[4] $P(Z = 0) \approx e^{-2.5} \approx 0.082$

[5] $m_Z(\theta) = \exp[\lambda(e^{\theta} - 1)] \;=\; \exp[2.5(e^{\theta} - 1)]$

$\log_{10} P(Z \geq 10) \leq -10\log_{10} 4 + 7.5\log_{10} e \approx -2.76$

理工２

[1] 決定係数は $R^2 \approx 0.639$ であり，自由度調整済み決定係数は $R^{*2} \approx 0.603$ となる。

[2] $F = 437.4/24.66 \approx 17.737$ であり，平方根 $t = \sqrt{17.737} \approx 4.21$ と自由度 10 の t 分布表より，有意水準 1%で有意である。

[3] $F = 154.0/27.75 \approx 5.55$ となる。F 分布表より検定は有意水準 5%で有意である。

[4] 完成させた分散分析表は以下である。

要因	平方和	自由度	平均平方	F 値
①	437.4	1	437.4	15.76
②	24.6	2	12.3	0.443
③	222.0	8	27.75	
計	684.0	11		

[5] ① については有意水準 1%で有意となり，② については有意水準 5%の検定でも有意にはならない。したがって，各設定温度における加工時間の平均値は，設定温度と直線的な関係でないとは言えない。

理工３

[1] $\beta_1 x_1$ と $\beta_{11} x_1^2$ が主効果を表す項であり，交互作用を表す項は $\beta_{12} x_1 x_2$ となる。

[2] 完成させた分散分析表は以下のようである。

要因	自由度	平方和	平均平方	F 値
接着 2 次モデル	9	258.35	28.71	15.43
誤差	10	18.60	1.86	
全体	19	276.95		

[3] $x_1 = 1$,　$x_2 = 1$,　$x_3 = 1$

[4] デザイン行列の列の直交性から推定値の不変性が判断できる。

[5] $\beta_{123}x_1x_2x_3, \beta_{112}x_1^2x_2, \beta_{122}x_1x_2^2, \beta_{113}x_1^2x_3, \beta_{133}x_1x_3^2, \beta_{223}x_2^2x_3, \beta_{233}x_2x_3^2$

理工４

[1] $x_0 = 3/5$ で，このとき $c_0 = f(3/5) = (6/5)^4 = 2.0736$ となる。

[2] $P(c_0 U \le f(Y)) = \dfrac{1}{c_0} = \left(\dfrac{5}{6}\right)^4 \approx 0.482$

[3] $h(x \mid X \ge 0) = f(x)$

[4] 次の手順に従う。(1) 区間 (0, 1) 上の独立な一様乱数 Y, U を 1 つずつ発生させ，(2) $c_0 U \le f(Y)$ であればその Y を X とし，そうでなければ (Y, U) は棄却し，(1) に戻る。乱数が 1 つ得られるまでの期待回数は 4.14 となる。

[5] c_0 より大きな任意の c_1 を用いると，必要な一様乱数の個数が増える。c_0 より小さな c_2 を用いると，$f(x) > c_2$ となる x に対応した範囲の乱数が得られる確率が小さくなる。

医薬 (22) 略解

医薬１

[1] 標本 1 と 2 のハザード比は $\dfrac{\lambda(t;x=1)}{\lambda(t;x=0)}=\exp(\beta)$ となり，比例ハザード性を満たす。

[2] $S_0(t)=\exp\left(-\int_0^t \lambda_0(s)ds\right)$

[3] $S(t\,;x)$ の両辺の 2 重対数を取ると $\log(-\log(S(t;x)))=\beta x+\log(-\log(S_0(t)))$ となり，$\log(-\log)$ 尺度上においてプロットが層間で平行になる。

[4] ハザード比の推定値は $\exp(0.06)=1.0618$ であり，ハザード比の 95%信頼区間は $(\exp(-0.29), \exp(0.41))\approx(0.75, 1.51)$ となる。

[5] ハザード比の 95%信頼区間は 1 を含むので，試験群とプラセボ群で差があるとはいえない。

医薬２

[1] 運転者のリスク：$1104/1418135\approx 0.0007$ (0.07%)
同乗者のリスク：$3074/1418135\approx 0.0021$ (0.21%)

[2] リスク比：$RR_c\approx 2.78$，オッズ比：$OR_c\approx 2.79$

[3] リスク比：$RR_{MH}=3074/1104\approx 2.78$，オッズ比：$OR_{MH}=2632/662\approx 3.98$

[4] 運転者群と同乗者群の交通事故の状況の分布は同じであるので，交絡バイアスは存在しない。リスク比は層間で共通の場合は集団全体のリスク比に一致するが，オッズ比は層間で共通であっても一般に集団全体のオッズ比に一致しない。

[5] 運転者では，シートベルトありに比べてシートベルトなしで死亡リスクが $2.78/1.01\approx 2.75$ 倍高くなることが示唆される。

医薬３

[1] $V[\boldsymbol{Y}]=ZGZ^T+R$

[2]

[2-1] $\hat{\boldsymbol{\beta}}=(X^TX)^{-1}X^T\boldsymbol{Y}$

[2-2] $E[\hat{\boldsymbol{\beta}}]=\boldsymbol{\beta}$

[2-3] $\hat{\boldsymbol{\beta}}=(20.17,-8.45)^T$

[3]

[3-1]

$$
Z=\begin{pmatrix}1&0&0&0\\1&0&0&0\\1&0&0&0\\0&1&0&0\\0&1&0&0\\0&1&0&0\\0&0&1&0\\0&0&1&0\\0&0&1&0\\0&0&0&1\\0&0&0&1\\0&0&0&1\end{pmatrix},\quad G=\begin{pmatrix}\sigma_1&0&0&0\\0&\sigma_2&0&0\\0&0&\sigma_3&0\\0&0&0&\sigma_4\end{pmatrix}
$$

[3-2] 表 1 の被験者内のデータには相関があると考えるのが自然である。[2] のモデルではすべてのデータが独立であることを仮定した解析となる。一方，[3-1] のモデルでは，被験者内のデータの相関を考慮した解析となる。

医薬４

[1] 尤度関数は $L=\binom{n}{y}\pi^y(1-\pi)^{n-y}$ であり，最尤推定量は $\hat{\pi}=y/n$ となる。

[2] 事後分布は $p(\pi\mid y)=\dfrac{\pi^y(1-\pi)^{n-y}}{B(y+1,n-y+1)}$ である。

[3] 最頻値は y/n となり，最尤推定量に一致する。

[4] 事後分布の期待値は $E[\pi\mid y]=\dfrac{y+1}{n+2}$ となる。$E[\pi\mid y]=\dfrac{n\times(y/n)+2\times(1/2)}{n+2}$ であるので，最尤推定量 $\hat{\pi}$ と事前分布の期待値 1/2 の重み付き平均である。

[5] Y の周辺分布は，確率関数が $p(y)=1/(n+1)$ $(y=0,1,\ldots,n)$ の離散一様分布である。

[6] $Y=4$ が最小値となる。

共通 (22) 略解

共通 1

[1] $E[D] = 0$, $V[D] = 72$。 $P(D \leq -4) \approx 0.32$ 。

[2] $E[Y \mid X = 132] = 129$, $V[Y \mid X = 132] = 63$。平均への回帰とは，2 回の測定 (X, Y) において，1 回目の測定値 x が平均 $\mu_X = E[X]$ よりも大きい（小さい）個体の 2 回目の測定値の期待値 $E[Y \mid X = x]$ は $\mu_Y = E[Y]$ よりも大きい（小さい）ものの，差の絶対値 $|E[Y \mid X = x] - \mu_Y|$ は $|x - \mu_X|$ よりも小さいことをいう。平均への回帰分は 3mmHg とみなされる。

[3] $V[X] = \tau^2 + \psi^2$， $V[Y] = \tau^2 + \psi^2$， $Cov[X, Y] = V[\theta] = \tau^2$。

$\tau^2 = 108$, $\psi^2 = 6^2$。

[4] $E[\theta \mid X = 132] = \mu + \beta(x - \mu)$, $V[\theta \mid X = 132] = \tau^2 - (\tau^2)^2/\sigma^2$ であり，数値を代入すると $E[\theta \mid X = 132] = 129$, $V[\theta \mid X = 132] = 27$ となる.

[5] $X = 132$ の人の 2 回目の測定値 Y の条件付き期待値が全平均に近づく理由は，1 回目の測定値が $X = x\ (> \mu)$ となった人の真値 θ の分布は x よりも小さな平均を持つ分布に従い，2 回目の測定値はその θ の分布を反映しているためと言える。
$P(Y \leq 128 \mid X = 132) \approx 0.45$。