2017.11.26

統計検定１級（統計応用）

解答と解説（人文）略解

人文１

[1] バリマックス回転は直交回転で，因子負荷量の 2 乗和が共通性に一致することから，A がバリマックス回転，B がプロマックス回転後の因子負荷量行列である。

[2] バリマックス回転は直交回転であるので，回転後の因子間相関係数は 0 である。

[3] プロマックス回転後の因子間相関係数は $r = 0.5$ である。

[4] $x = \pm 0.2$，$y = 0.2$ もしくは $y = -0.8$ となる。

[5] 因子分析モデル $\boldsymbol{x} = \Lambda \boldsymbol{f} + \boldsymbol{\xi}$ では，Λ に対し右から正則行列 T をかけるだけの自由度を有する。その正則行列 T が直交行列の場合，直交回転といい，直交行列でないとき斜交回転という。因子負荷量行列の解釈が容易になるような適切な回転をすることが望ましい。

人文２

[1]　各平方和は

$$SST=\sum_{i=1}^{n_A}(y_{Ai}-\overline{y})^2+\sum_{i=1}^{n_B}(y_{Bi}-\overline{y})^2+\sum_{i=1}^{n_C}(y_{Ci}-\overline{y})^2$$

$$SSB=n_A(\overline{y}_A-\overline{y})^2+n_B(\overline{y}_B-\overline{y})^2+n_C(\overline{y}_C-\overline{y})^2$$

$$SSW=\sum_{i=1}^{n_A}(y_{Ai}-\overline{y}_A)^2+\sum_{i=1}^{n_B}(y_{Bi}-\overline{y}_B)^2+\sum_{i=1}^{n_C}(y_{Ci}-\overline{y}_C)^2$$

であり $SST=SSB+SSW$ の関係がある。

[2]　$df_T=n-1$，$df_B=2$，$df_W=n-3$ である。

[3]　対立仮説は「$H_1:\mu_A=\mu_B=\mu_C$ でない」である。検定統計量は $F=\dfrac{SSB/2}{SSW/(n-3)}$ であり，H_0 の下で自由度 $(2,n-3)$ の F 分布に従う。

[4]　検定統計量の値は $F=\dfrac{1290/2}{1736/12}\approx 4.46$ である。検定は有意水準 5%で有意であり，各母平均が等しいという帰無仮説は棄却される。

[5]　$\overline{y}_{BC}=(n_B\overline{y}_B+n_C\overline{y}_C)/(n_B+n_C)$ としたとき，統計量 $t=\dfrac{\overline{y}_A-\overline{y}_{BC}}{\sqrt{\left(\dfrac{1}{n_A}+\dfrac{1}{n_B+n_C}\right)\dfrac{SSW}{n-3}}}$ は H_0 の下で自由度 $n-3$ の t 分布に従うことを用いて検定する。数値例では，$\overline{y}_{BC}=64.5$ であり，$t\approx -2.96$ であるので，有意水準 1%で有意である。

人文３

[1]　$\alpha = 20$ および $\beta = 0.6$ である。

[2]　$P(Y \geq 60 \mid X = 60) = P\left(Z \geq \dfrac{60 - 56}{12} \right) = P(Z \geq 1/3) \approx 0.369$　となる。

[3]　$\xi = 50 + 15\nu \approx 50 + 15 \times 0.8 = 62$ となる。

[4]　$\eta = E[Y \mid X \geq 50] = \alpha + \beta\,\xi = 20 + 0.6 \times 62 = 57.2$ となる。

[5]　$\xi = 62 > \eta = 57.2$ である。平均への回帰現象が観察される。

人文４

[1]　$V[\bar{y}] = \dfrac{\sigma^2}{n}$ となる。

[2]　$E[\bar{y}] = \dfrac{1}{N}\sum_{\ell=1}^{L} E[\hat{T}_\ell] = \dfrac{1}{N}\sum_{\ell=1}^{L} T_\ell = \dfrac{T}{N} = \mu$ より不偏性が示される。分散は $V[\bar{y}] = \dfrac{1}{n}\sum_{\ell=1}^{L} \dfrac{N_\ell}{N}\sigma_\ell^2$ となる。

[3]　層の間で平均 μ_ℓ が大きく異なるように層化するのが望ましい。

[4]　S さんの案では $V[\bar{y}] = 1$ となり，T さんの案で $V[\bar{y}] = 376/360 \approx 1.044$ となって，S さんの案のほうがサンプルサイズは少なく分散は小さいので優れている。しかし，標本抽出の手間が T さんの提案のほうが小さく，二人の提案は甲乙つけがたい。

人文５

[1]　$E[X]=5\times\frac{1}{2}=\frac{5}{2}$，$V[X]=5\times\frac{1}{2}\times\frac{1}{2}=\frac{5}{4}$ となる。

[2]　$E[Y]=\frac{5}{2}$，$V[Y]=\frac{5}{12}$ である。

[3]　$E[X]=E[Y]$ であり，$V[X]>V[Y]$ となる。

[4]　$L_A=8.0$ となる。また，$\frac{L_A}{L_B}=\sqrt{\frac{13}{12}}$ である。

[5]　$n=204$ が n の最小値を与える。

2017.11.26

解答と解説（社会）略解

社会 1

[1] $t = 0.0827/0.00757 = 10.9247$ となり，有意水準 5%で有意であり帰無仮説は棄却される。

[2] 95%信頼区間は (0.0679, 0.0975) となる。

[3] $0.0827 \times 2 = 0.1654$ であるので，16.54%上昇すると期待される。

[4] 誤差項 ε は観測されない「能力」を含むと考えられ，「能力」は「学歴」と正の相関があると予想できるので，β_1 の最小二乗推定量は正のバイアスをもつと予想される。

[5] 操作変数とは，回帰モデルにおける誤差項と相関を持たず，説明変数に影響を与え，目的変数には説明変数を通じてのみ影響を与える変数のことを言う。「父親の学歴」はその条件をおおむね満たす。

[6] 「父親の学歴」と「母親の学歴」双方とも「能力」と独立で，かつ「能力」を「父親の学歴」,「母親の学歴」に回帰したとき，回帰係数の少なくとも一方が 0 でないことが条件。これらの条件はおおむね満たされる。

社会２

[1] t 値 $= -6.58$ であり，有意水準 5%で帰無仮説は棄却される。

[2] t 値 $= 3.45$ で，有意水準 5%で帰無仮説は棄却される。

[3] t 値 $= 0.963$ で，有意水準 5%で帰無仮説は棄却されない。

[4] 回帰式より 3.16%進学率が上昇すると予測される。

[5] モデル 1 はトレンドを無視して回帰しているため，「みかけの回帰」が発生していると考えられる。一方，モデル 2 あるいは 3 はトレンドを除去して回帰分析しているので，「みかけの回帰」が発生しないと考えられ，合理的な結果となっている。

[6] y_t，x_t のそれぞれに Dicky-Fuller 検定，すなわち $H_0 : \rho = 1$ を対立仮説 $H_1 : \rho < 1$ に対して検定し，帰無仮説を棄却できなければモデル 3 を選択し，従属変数および独立変数の両変数に対して帰無仮説を棄却できればモデル 2 を選択する。

社会３

[1] $E[X]=V[X]=\lambda$ である。

[2] 対数尤度関数 $l(\lambda)=\log\lambda\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\log x_i!-n\lambda$ の微分により，λ の最尤推定値は $\hat{\lambda}=\bar{x}$ と観測データの標本平均となる。

[3] $\xi=\dfrac{\lambda}{1-e^{-\lambda}}$，$V[X\mid X\geq 1]=\dfrac{\lambda-\lambda(\lambda+1)e^{-\lambda}}{(1-e^{-\lambda})^2}$ と求められる。λ の推定値としては正の偏りを持つ。

[4] $E[X]=E[Y+1]=\lambda+1$，分散は $V[X]=V[Y]=\lambda$ である。

社会４

[1]　$V[\bar{y}] = \dfrac{\sigma^2}{n}$ となる。

[2]　$E[\bar{y}] = \dfrac{1}{N}\sum_{\ell=1}^{L} E[\hat{T}_\ell] = \dfrac{1}{N}\sum_{\ell=1}^{L} T_\ell = \dfrac{T}{N} = \mu$ より不偏性が示される。分散は $V[\bar{y}] = \dfrac{1}{n}\sum_{\ell=1}^{L}\dfrac{N_\ell}{N}\sigma_\ell^2$ となる。

[3]　層の間で平均 μ_ℓ が大きく異なるように層化するのが望ましい。

[4]　S さんの案では $V[\bar{y}] = 1$ となり，T さんの案で $V[\bar{y}] = 376/360 \approx 1.044$ となって，S さんの案のほうがサンプルサイズは少なく分散は小さいので優れている。しかし，標本抽出の手間が T さんの提案のほうが小さく，二人の提案は甲乙つけがたい。

社会５

[1]　$E[X]=5\times\frac{1}{2}=\frac{5}{2}$，$V[X]=5\times\frac{1}{2}\times\frac{1}{2}=\frac{5}{4}$ となる。

[2]　$E[Y]=\frac{5}{2}$，$V[Y]=\frac{5}{12}$ である。

[3]　$E[X]=E[Y]$ であり，$V[X]>V[Y]$ となる。

[4]　$L_A=8.0$ となる。また，$\frac{L_A}{L_B}=\sqrt{\frac{13}{12}}$ である。

[5]　$n=204$ が n の最小値を与える。

2017.11.26

解答と解説（理工）略解

理工１

[1]　期待値は $E[X]=\int_0^{\infty} x\cdot\frac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}dx=\alpha\beta$ と求められる。また，$E[X^2]=\alpha(\alpha+1)\beta^2$ より，$V[X]=E[X^2]-(E[X])^2=\alpha\beta^2$ となる。

[2]　対数尤度関数は $l=-n\alpha\log\beta-n\log\Gamma(\alpha)+(\alpha-1)\sum_{i=1}^{n}\log x_i-\frac{1}{\beta}\sum_{i=1}^{n}x_i$ となる。

[3]　対数尤度関数を各パラメータで偏微分して 0 と置き，$\beta=\bar{x}_n/\alpha$ および $\psi(\alpha)-\log\alpha=\log\frac{\tilde{x}_n}{\bar{x}_n}$ を得る。

[4]　$\hat{\alpha}=1$ および $\hat{\beta}=\tilde{x}_{10}/\hat{\alpha}=1000$ を得る。

[5]　$\eta(\alpha)$ は α についての狭義単調増加関数であることが示され，$\eta(\alpha)=A(\boldsymbol{x})$ は α について必ず一つの解を持つ。この α についての解を尤度方程式から導かれる式に代入すれば β の解が唯一得られる。

理工２

[1] X の期待値は $E[X]=\dfrac{1}{\lambda}$ であり，$x>0$ において，X の累積分布関数は $F(x)=1-e^{-\lambda x}$ であるので，$\dfrac{f(x)}{1-F(x)}=\lambda$ すなわち瞬間故障率は x によらず一定値 λ となる。

[2] 任意の $x\geq 0$ において $P(X_1>x)=P(N(t)=0)=e^{-\lambda x}$ である。よって，X_1 の累積分布関数は $G(x)=1-e^{-\lambda x}$ となり，パラメータ λ の指数分布の累積分布関数に一致する。

[3] MTTF の 95%下側限界値は $\mathrm{MTTF}\geq -t_0/\log 0.05\approx t_0/3$ となる。時刻 t_0 で故障が観測されていなかった場合， 95%下側限界値は $t_0/3$ よりも大きいとしかいえない。

[4] 要求される条件は $P(X_1>1000)\geq 0.999$ であるので，$\lambda\leq -\dfrac{\log 0.999}{1000}\approx 10^{-6}$ を得る。

理工３

[1]　T の期待値は $\xi = E[T \mid T > t] = \mu + t$ となる。

[2]　対数尤度関数は $l_1(\mu) = -2\log\mu - \dfrac{t_1 + t_2}{\mu}$ となる。μ の最尤推定値は $\hat{\mu} = \dfrac{t_1 + t_2}{2}$ となる。

[3]　対数尤度関数は $l_2(\mu) = -\log\mu - \dfrac{t_1 + t}{\mu}$ であり，μ の最尤推定値は $\tilde{\mu} = t_1 + t$ となる。

[4]　$\mu^{(k+1)} = \dfrac{t_1 + \mu^{(k)} + t}{2} = \dfrac{t_1 + t}{2} + \dfrac{\mu^{(k)}}{2}$ となる。

[5]　$\mu^{(k+1)} = \left(1 + \dfrac{1}{2} + \dfrac{1}{2^2} + \cdots + \dfrac{1}{2^k}\right)\dfrac{t_1 + t}{2} + \dfrac{\mu^{(0)}}{2^k}$ で $k \to \infty$ とすると，カッコ内は $1/(1-1/2) = 2$ に収束し，第 2 項は 0 に収束する。よって，$k \to \infty$ の極限では，初期値 $\mu^{(0)}$ によらず $\mu^{(k+1)} \to t_1 + t$ となる。

理工４

[1]　$E[X]=\frac{1}{\mathrm{B}(a,b)}\int_0^1 x\cdot x^{a-1}(1-x)^{b-1}dx=\frac{a}{a+b}$ 。問題文の確率分布は $Beta(2,3)$ であり，$\theta=$ $2/(2+3)=0.4$ である。

[2]　$f(x)$ を微分して 0 と置き $x_{\max}=1/3$ を得る。このとき $f_{\max}=\frac{16}{9}$ である。

[3]　いくつかの方法がある。

(1) 棄却法：区間 (0, 1) 上の一様乱数 U を生成し，[2] で求めた $f_{\max}$ を用いて，確率 $f(U)/f_{\max}$ で U を保持し，確率 $1-f(U)/f_{\max}$ で棄却する。

(2) 逆関数法：$Beta(a,b)$ の累積分布関数 $F(x)$ の逆関数 F^{-1} が求まれば，区間 (0, 1) 上の一様乱数を U として，$X=F^{-1}(U)$ とする。

(3) $Beta(a,b)$ の特性を利用する方法：Y，Z をそれぞれ形状パラメータ a，b をもち，共通の尺度パラメータをもつガンマ分布に従う乱数としたとき，$X=Y/(Y+Z)$ は $Beta(a,b)$ に従うことを利用する。

[4]　$V[\hat{\theta}_1]=\frac{1}{n}\left(\frac{1}{5}-\frac{4}{25}\right)=\frac{1}{n}\times\frac{1}{25}$，$V[\hat{\theta}_2]=\frac{1}{n}\left(\frac{8}{35}-\frac{4}{25}\right)=\frac{1}{n}\times\frac{12}{175}\approx\frac{1}{n}\times\frac{1.71}{25}$ を得る。

[5]　方法 (ii) による推定値の分散は方法 (i) による推定値に比べ分散が 1.71 倍で，同じ精度を得るためには，シミュレーション回数を 1.71 倍にする必要がある。しかし，方法 (i) は方法 (ii) に比べて乱数を生成する手間が余分にかかるという難点がある。逆に方法 (ii) は関数 $uf(u)$ の計算に多少時間がかかる。

理工５

[1]　$E[X]=5\times\frac{1}{2}=\frac{5}{2}$，$V[X]=5\times\frac{1}{2}\times\frac{1}{2}=\frac{5}{4}$　となる。

[2]　$E[Y]=\frac{5}{2}$，$V[Y]=\frac{5}{12}$　である。

[3]　$E[X]=E[Y]$　であり，$V[X]>V[Y]$　となる。

[4]　$L_A=8.0$　となる。また，$\frac{L_A}{L_B}=\sqrt{\frac{13}{12}}$　である。

[5]　$n=204$　が　n　の最小値を与える。

2017.11.26

<u>解答と解説（医薬）略解</u>

<u>医薬１</u>

[1]　条件付き確率の定義より簡単に導かれる。

[2]　$f(u) = -S'(u)$ であること，および $S(0) = 1$ であることに注意して容易に導かれる。

[3]　$F(T)$ は区間 $(0, 1)$ 上の一様分布に従うことから導かれる。

[4]　Cox 比例ハザードモデルより $\Lambda(t \mid \boldsymbol{Z}) = \exp[\boldsymbol{\beta}'\boldsymbol{Z}]\Lambda_0(t)$ となり，$-\log S(t \mid \boldsymbol{Z}) = \Lambda_0(t)\exp[\boldsymbol{\beta}'\boldsymbol{Z}]$ の対数を取ることにより導かれる。$h(t) = \log \Lambda_0(t)$ である。

[5]　上問 [4] の解答より $\varepsilon(t) = \log(-\log(S(t \mid \boldsymbol{Z})))$ であり，T を確率積分変換した $F(T \mid \boldsymbol{Z})$ が区間 $(0, 1)$ 上の一様分布に従うことから導かれる。

[6]　次の手順で T に対する乱数を得ることができる。

Step 1：区間 $(0, 1)$ 上の一様分布に従う乱数 U を生成し，$Z = 40 + 30U$ とする。

Step 2：U とは独立に区間 $(0, 1)$ 上の一様分布に従う乱数 V を生成し，$\varepsilon(T) = \log(-\log V)$ とする。

Step 3：$T = \exp\left[-\frac{Z}{3} + \frac{2}{3}\log(-\log V)\right]$ とする。

医薬２

[1]　$\alpha = 0.042$ となる。

[2]　$P(Z(0.4) \geq z_1 \mid H_0) = 0.008$ より $z_1 = 2.41$ を得る。

[3]　$Cov[Z(0.4), Z(1)] = \sqrt{0.4}$ となる。

[4]　$\int_{-\infty}^{z_1}\int_{z_2}^{\infty} \phi_2(x_1, x_2 : \sqrt{0.4})dx_1 dx_2 = 0.042$ を満たす値となる。

[5]　最終解析での α は $0.05 - 0.05 \times (0.7)^2 = 0.0255$ となる。

医薬３

[1]　$P(X_1 \le r_1 \mid \pi = \pi_0) = \sum_{j=0}^{r_1} {}_{n_1}C_j \pi_0{}^j (1-\pi_0)^{n_1-j}$　である。

[2]　$B(x;\pi,n) = P(X \le x) = \sum_{k=0}^{x} b(k;\pi,n)$　として，

$$\beta = \sum_{x=0}^{r_1} b(x;\pi_1,n_1) + \sum_{x=r_1+1}^{\min(n_1,r)} b(x;\pi_1,n_1) B(r-x;\pi_1,n_2)$$

となる。

[3]　$\alpha = \sum_{x=r_1+1}^{n_1} b(x;\pi_0,n_1)\{1 - B(r-x;\pi_0,n_2)\}$　を得る。

[4]　$\alpha = 0.024$，　$\beta = 0.089$

[5]　$E[N \mid \pi = \pi_0] = 7.9$　となる。

医薬４

[1]　$\psi_1 = 7.71$，$\psi_2 = 6.03$，$\psi_1 = 3.61$ となる。

[2]　各オッズ比はそれぞれ，31.92，6.87，3.87 である。

[3]　対数オッズ比の重み付き平均は $\bar{Y} = 1.827$ で，共通オッズ比の推定値は $\hat{\psi} = 6.22$ となる。重み付き平均の分散は $V[\bar{Y}] = \dfrac{1}{26.167}$ と求められる。

[4]　共通対数オッズ比 $\log\psi$ の 95%信頼区間の上下限は $\log\psi_L = 1.445$，$\log\psi_U = 2.209$ であり，共通オッズ比の 95%信頼区間の上下限は $\hat{\psi}_L = 4.24$，$\hat{\psi}_U = 9.11$ となる。

[5]　正規分布の母平均の均一性の検定あるいは Breslow-Day 検定を適用する。

医薬５

[1]　$E[X]=5\times\dfrac{1}{2}=\dfrac{5}{2}$，$V[X]=5\times\dfrac{1}{2}\times\dfrac{1}{2}=\dfrac{5}{4}$ となる。

[2]　$E[Y]=\dfrac{5}{2}$，$V[Y]=\dfrac{5}{12}$ である。

[3]　$E[X]=E[Y]$ であり，$V[X]>V[Y]$ となる。

[4]　$L_A=8.0$ となる。また，$\dfrac{L_A}{L_B}=\sqrt{\dfrac{13}{12}}$ である。

[5]　$n=204$ が n の最小値を与える。