2016.11.27

統計検定１級（統計応用）

解答と解説（人文）略解

人文１

[1] X_1 の共通性は 0.84 である。

[2] X_3 と因子 1 の間の相関係数は 0.7 である。

[3] X_3 と X_5 の相関係数は 0.58 となる。

[4] 直交回転の場合は回転後の因子は互いに無相関である。これに対し，斜交回転の場合は因子間に相関があってもよい。

人文２

[1] 大学全体での平均値は，2011 年度が 470.5，2014 年度が 468.5 である。両コースでは平均点の伸びが観測されているが，上級コースの履修者が減少し，一般コースの履修者が増加したため大学全体では平均点が下がっている。

[2] 上級コースの割合は $P(X \geq 500) \approx 0.3446$ であり，一般コースの割合は $1 - 0.3446 = 0.6554$ である。

[3] 上級コースの点数の平均は 553.50 となり，一般コースの平均は 423.05 となる。

人文 3

[1] $R_s = 0, 1$ だから $R_s^2 = R_s$，$\sum_s R_s = 1$ だから $s \neq t$ のとき $R_s R_t = 0$ となることを用いる。

[2] $E(\bar{x}_s) = \mu$, $\mathrm{var}(\bar{x}_s) = \sigma_s/n$ である。前問と同様に，$x_{TS} = \sum_s R_s \bar{x}_s$ と表されることを用いる。

[3] 分散は以下の通りで，$V_{TS} \geq V_{SI}$ かつ等号が成立するのは $\mu_s = \mu$ となる場合である。

$$V_{SI} = \frac{1}{mn}\sigma^2 = \frac{1}{mn}\sum_s W_s \sigma_s^2 + \frac{1}{mn}\sum_s W_s(\mu_s - \mu)^2$$

$$V_{TS} = \frac{1}{m}\mathrm{var}(x_{TS}) = \frac{1}{mn}\sum_s W_s \sigma_s^2 + \frac{1}{m}\sum_s W_s(\mu_s - \mu)^2$$

[4] 一定の予算で調査を実施する場合には，単純無作為抽出より 2 段抽出の方が精度が高くなることがある。

人文４

[1] モデルにおける推定すべき母数は 9 個である。

[2] 自由度は $10 - 9 = 1$ である。

[3] $x_1 = a_1 f_1 + e_1$ となる。x_1 は平均偏差化されているので，切片は必要ない。

[4] 母数で表現された共分散行列は以下のようになる。

分散共分散行列

	x_1	x_2	x_3	x_4
x_1	$a_1^2 + ve_1$	-	-	-
x_2	a_1a_2	$a_2^2 + ve_2$	-	-
x_3	a_1a_3r	a_2a_3r	$a_3^2 + ve_3$	-
x_4	a_1a_4r	a_2a_4r	a_3a_4	$a_4^2 + ve_4$

(5) 因子間の共分散に $r = 0$ という制約を置くと，このモデルは識別されない。因子間の共分散 $r = 0$ という制約を置くと推定値の一意性がなくなる。

人文５

[1] 母平均 μ_B の信頼係数 95%の信頼区間は (14.46, 25.54) と求められる。

[2] 2 標本 t 検定の検定統計量の値は $t^* = 2.117$ であり，検定は有意水準 5%で有意となる．母平均の差の信頼係数 95%の信頼区間は (0.23, 13.77) となる．

[3] 重なり具合が

$$D < \left\{ 1 - \frac{\sqrt{2}}{2} \frac{t_{0.025}(2n-2)}{t_{0.025}(n-1)} \right\} \times 2t_{0.025}(n-1)\frac{s}{\sqrt{n}}$$

であれば，母平均の差の検定は有意となる．中括弧の中は $n \to \infty$ のとき $1-\sqrt{2}/2 \approx 0.293$ に近づく．

2016.11.27

解答と解説（社会）略解

社会 1

[1] 95%信頼区間は $-1.77 \leq \beta_1 \leq 2.28$ である。

[2] β_1, β_2 は，正となることが期待される。

[3] 仮説 $H_0 : \beta_1 + \beta_2 = 1$ に関する統計量は $t = -0.582$ となり，仮説は棄却されない。

[4] 95%信頼区間は $0.555 \leq \beta_1 \leq 0.919$ である。

[5] 二つの説明変数に強い線形関係が存在している，典型的な（多重）共線性の状況である。

社会２

[1] 幾何平均と算術平均に関する不等式 $(1/n)\sum_i x_i \geq (\prod_i x_i)^{1/n}$ から明らかである。

[2] Dutot 指数と Jevons 指数は時点の対称性を満たすが，Carli 指数は満たさない。

[3] $P_P = (\sum_i p_{ti}q_{ti})/(\sum_i p_{0i}q_{ti})$，$Q_L = (\sum_i p_{0i}q_{ti})/(\sum_i p_{0i}q_{0i})$，$Q_P = (\sum_i p_{ti}q_{ti})/(\sum_i p_{ti}q_{0i})$

[4] (4-1), (4-2), (4-3) は，いずれも $x_i = p_{ti}/p_{0i}, y_i = q_{ti}/q_{0i}$，$V_t = \sum_i p_{ti}q_{ti}$，$w_i = v_{0i}/V_0$ を利用すればよい。
(4-5) 価格と購入数量の相関係数は負となることが期待され，$P_L \geq P_P$ となる傾向がある。

社会3

[1] $\bar{y}_i$ と y_{it} との差を取ると，$y_{it} - \bar{y}_i = (x'_{it} - \bar{x}'_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$ となり，固別効果 α_i が消去される。$\hat{\boldsymbol{\beta}}_0$ は，この nT 個の式に最小二乗法を適用して得られる。

[2] $\hat{\alpha}_j = 2.5$

[3] $\mathrm{var}(u_i) = \sigma_\alpha^2 \boldsymbol{\iota}_T \boldsymbol{\iota}'_T + \sigma_\epsilon^2 I_T$

[4] 回帰係数 $\boldsymbol{\beta}$ の GLS 推定量 $\hat{\boldsymbol{\beta}}_1$ はこの回帰式に GLS 法を適用して得られるものである。通常は σ_α^2 と σ_ε^2 は未知だから，適当な一致推定量を代入する。$\hat{\Omega} = \hat{\sigma}_\alpha^2 \boldsymbol{\iota}_T \boldsymbol{\iota}'_T + \hat{\sigma}_\epsilon^2 I_T$ とすると，

$\hat{\boldsymbol{\beta}}_1 = (\sum_{i=1}^n X'_i \hat{\Omega}^{-1} X_i)^{-1} (\sum_{i=1}^n X'_i \hat{\Omega}^{-1} y_i)$

[5] H_0 が成り立たなければ $\hat{\boldsymbol{\beta}}_1$ は一致性を持たない。$\hat{\beta}_0$ は仮定 H_0 に関わらず一致性を持つ。

[6] 検定統計量 $\hat{q}' \left[\mathrm{var}(\hat{\boldsymbol{\beta}}_0) - \mathrm{var}(\hat{\boldsymbol{\beta}}_1)\right]^{-1} \hat{q}$ は H_0 の下では漸近的に自由度 p のカイ二乗分布に従う。

社会４

[1] μ の推定値は 12.0 であり，標準誤差は $SE[\overline{X}]=\dfrac{12.0}{\sqrt{10}}\approx 3.795$ となる。

[2] n 個中観測が m 個あり，打ち切りが $n-m$ 個あるとすると，μ の最尤推定値は $\tilde{\mu}=\dfrac{1}{m}\left\{\sum_{i=1}^{m}x_i+(n-m)c\right\}$ で与えられる。$n=12$，$m=2$ では $\tilde{\mu}=\dfrac{1}{10}(120+60)=18.0$ となる。

[3] m 人分のデータが観測されたときの μ の最尤推定値は

$$\bar{x}=\breve{\mu}-\frac{ce^{-c/\breve{\mu}}}{1-e^{-c/\breve{\mu}}}$$

を満足する $\breve{\mu}$ として得られる。これより，

$$\mu^{*(t+1)}=\bar{x}+\frac{ce^{-c/\mu^{*(t)}}}{1-e^{-c/\mu^{*(t)}}}$$

という反復計算スキームが得られる。

社会５

[1] 母平均 μ_B の信頼係数 95%の信頼区間は (14.46, 25.54) と求められる。

[2] 2 標本 t 検定の検定統計量の値は $t^* = 2.117$ であり，検定は有意水準 5%で有意となる．母平均の差の信頼係数 95%の信頼区間は (0.23, 13.77) となる．

[3] 重なり具合が

$$D < \left\{ 1 - \frac{\sqrt{2}}{2} \frac{t_{0.025}(2n-2)}{t_{0.025}(n-1)} \right\} \times 2t_{0.025}(n-1)\frac{s}{\sqrt{n}}$$

であれば，母平均の差の検定は有意となる．中括弧の中は $n \to \infty$ のとき $1-\sqrt{2}/2 \approx 0.293$ に近づく．

2016.11.27

解答と解説（理工）略解

理工１

[1]　不偏性は容易に示される。$V[T_1] = V[T_2] = V[T_3] = 2\sigma^2$ である。

[2]　推定量は $T_4 = \left(Y_{13} - \dfrac{Y_{33}+Y_{43}}{2}\right) - \left(Y_{24} - \dfrac{Y_{34}+Y_{44}}{2}\right)$ となり，分散は $V[T_4] = 3\sigma^2$ である。

[3]　T の分子の各項はそれぞれ $\tau_1 - \tau_2$ の不偏推定量であるので, 全体としても不偏となる。

$V[T] = \dfrac{6+3w^2}{(3+w)^2}\sigma^2$ であり，これを最小にするのは $w = 2/3$ で，$V[T] = \dfrac{6}{11}\sigma^2$ となる。

[4]　T は最小分散線形不偏推定量である。

理工２

[1]　期待値は $E[Y_t]=0$ となる。

[2]　$Cov[Y_t, Y_{t-s}] =$ $E[Y_t Y_{t-s}] = \sigma^2 \sum_{j=s}^{\infty} g_j g_{j-s}$ を得る。

[3]　$C(1) = a_1 C(0) + a_2 C(1)$ である。

[4]　$C(2) = a_1 C(1) + a_2 C(0)$ である。

[5]　ユール・ウォーカー方程式は

$$\begin{pmatrix} C(0) & C(1) \\ C(1) & C(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} C(1) \\ C(2) \end{pmatrix}$$

となる。

理工３

[1]　正規分布は平均 μ を中心に一山型で左右対称であるので，μ について左右対称な区間がその中に入る確率を最大にする。

[2]　区間 $[\mu-2\sigma, \mu+2\sigma]$ の外側の確率が不良率で，数表より 0.0456 を得る。

[3]　自由度 $n-1$ のカイ二乗分布を用いて信頼区間が導出される。

[4]　工程能力指数は一般に 1.33 以上が好ましいという議論があり，C_P の点推定値は 2.0 であるので，この製品の製造プロセスは十分な管理状態にあるといえる。

[5]　$a = \dfrac{\sqrt{n-1}\,\Gamma((n-2)/2)}{\sqrt{2}\,\Gamma((n-1)/2)}$ となる。

理工４

[1] モーメント母関数は定義から求められ，$E[X]=V[X]=\lambda$ である。

[2] λ が整数でないときは，x が λ を超えない最大の整数のときに $p(x)$ の最大値を与える。λ が整数のときは $x=\lambda-1$ と $x=\lambda$ が最頻値となる。

[3] 最尤推定値は標本平均で与えられ，$\hat{\lambda}=1.82$ となる。

[4] Y のモーメント母関数は $M_Y(t)=\dfrac{e^{-\lambda}}{1-e^{-\lambda}}(\exp[\lambda e^t]-1)$ であり，期待値は $E[Y]=\dfrac{\lambda}{1-e^{-\lambda}}$ となる。関係式 $\bar{y}=\dfrac{\hat{\lambda}}{1-e^{-\hat{\lambda}}}$ が得られ，これを数値的に解けば最尤推定値 $\hat{\lambda}$ が求められる。

理工５

[1] 母平均 μ_B の信頼係数 95%の信頼区間は (14.46, 25.54) と求められる。

[2] 2 標本 t 検定の検定統計量の値は $t^* = 2.117$ であり，検定は有意水準 5%で有意となる．母平均の差の信頼係数 95%の信頼区間は (0.23, 13.77) となる．

[3] 重なり具合が

$$D < \left\{ 1 - \frac{\sqrt{2}}{2} \frac{t_{0.025}(2n-2)}{t_{0.025}(n-1)} \right\} \times 2 t_{0.025}(n-1) \frac{s}{\sqrt{n}}$$

であれば，母平均の差の検定は有意となる．中括弧の中は $n \to \infty$ のとき $1 - \sqrt{2}/2 \approx 0.293$ に近づく．

2016.11.27

解答と解説（医薬）略解

医薬１

[1] 有意水準 5%の両側検定で有意でない。帰無仮説は，差 z の母平均 μ に関して，H_0: $\mu = 0$ である。対応のある t 検定の妥当性の仮定は，差 z の母集団分布の正規性と観測値の独立性である。

[2] 符号検定が必要とする仮定はなく，符号付き順位検定は差 z の分布が中央値に関して左右対称であることを仮定し，両検定とも観測値が互いに独立である必要がある。2つの検定は共に差 z の中央値 θ に関して，H_0: $\theta = 0$ を検定する。

[3] $E[T \mid H_0] = np = n/2$，$V[T \mid H_0] = np(1-p) = n/4$ となる。

[4] 両側 P 値は

$$P(T \geq t^* \mid H_0) = \sum_{i=t^*}^{n} {}_nC_i (0.5)^n$$

の 2 倍となる。本データでは，両側 P 値 $= 0.125$ である。

[5] $E[W^+ \mid H_0] = \dfrac{n(n+1)}{4}$，$V[W^+ \mid H_0] = \dfrac{n(n+1)(2n+1)}{24}$ となる。正規近似に基づく検定統計量は $z = \dfrac{W^+ - E[W^+ \mid H_0]}{\sqrt{V[W^+ \mid H_0]}}$ で，表 1 の数値では $z = \dfrac{27-14}{\sqrt{35}} = 2.197$ となる。検定は有意水準両側 5%で有意であり，差 z の中央値は 0 でないといえる。

医薬２

[1]　母比率の一様性の検定に関する最も一般的な検定統計量は，

$$X=\frac{1}{p(1-p)}\sum_{i=1}^{n}n_i(p_i-p)^2$$

で，H_0 の下で近似的に自由度 3 のカイ二乗分布に従う。

[2]　この臨床試験のデータからは，4 群の間で母有効率に差があるとはいえない.

[3]　重み付き最小二乗法により $\hat{\beta}=\dfrac{\sum_i n_i(p_i-p)(x_i-\overline{x})}{\sum_i n_i(x_i-\overline{x})^2}$ が示され，標準誤差は

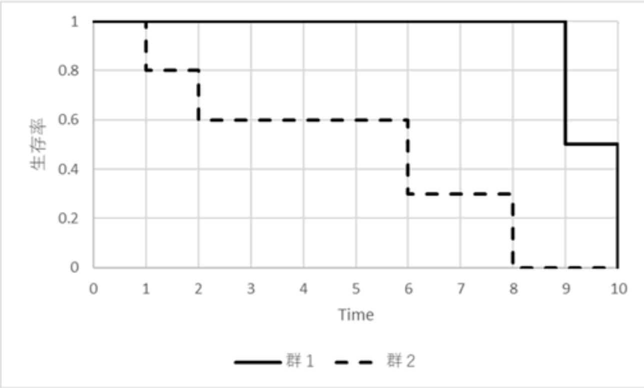
$SE[\hat{\beta}\,|\,H_0]=\sqrt{\dfrac{p(1-p)}{\sum_i n_i(x_i-\overline{x})^2}}$ と求められる。検定統計量は $Y=\dfrac{\hat{\beta}^2}{p(1-p)}\sum_i n_i(x_i-\overline{x})^2$

となる。

[4]　今回の臨床試験データでは，分割表の行変数に順序があるという情報を使用する傾向性検定を用いるべきで，それにより対立仮説の方向を規定しない検定よりも検出力が増す。

医薬３

[1] 各群でのカプラン・マイヤー曲線は以下のようである。



[2] ログランク検定のスコアは $u = \sum_t (d_t - n_{1t}/n_t)$ となり，分散は $v = \sum_t \frac{n_{1t}}{n_t}\frac{(n_t - n_{1t})}{n_t}$ となるので， $\chi^2 = u^2/v = 5.6225$ となり，平方根は 2.3712 であるので，標準正規分布の上側確率は 0.0089，P 値はその 2 倍の 0.0178 となる。

[3] 部分尤度は以下のようになる。

$$L = \frac{e^{\beta}}{5+5e^{\beta}} \times \frac{e^{\beta}}{5+4e^{\beta}} \times \frac{e^{\beta}}{3+2e^{\beta}} \times \frac{e^{\beta}}{2+e^{\beta}} \times \frac{1}{2} \times \frac{1}{1}$$

医薬４

[1]　X_i の中央値は $m = e^{\mu}$ となる。

[2]　X_i の累積分布関数を x で微分し，$(\log(x))' = 1/x$ であることより与式を得る。

[3]　$E[X_i] = M_Y(1) = \exp[\mu + \sigma^2/2]$，$V[X_i] = \exp[2\mu + \sigma^2](\exp[\sigma^2] - 1)$ と求められる。

[4]　$\sigma^2 = \log 2$ である。

[5]　AUC の中央値の比の信頼区間は (0.83, 1.48) となる。この信頼区間は同等域 (0.8, 1.25) に含まれないことから，2 つの薬剤の生物学的同等性は成り立つとは言えない。

医薬５

[1] 母平均 μ_B の信頼係数 95%の信頼区間は (14.46, 25.54) と求められる。

[2] 2 標本 t 検定の検定統計量の値は $t^* = 2.117$ であり，検定は有意水準 5%で有意となる．母平均の差の信頼係数 95%の信頼区間は (0.23, 13.77) となる．

[3] 重なり具合が

$$D < \left\{ 1 - \frac{\sqrt{2}}{2} \frac{t_{0.025}(2n-2)}{t_{0.025}(n-1)} \right\} \times 2t_{0.025}(n-1) \frac{s}{\sqrt{n}}$$

であれば，母平均の差の検定は有意となる．中括弧の中は $n \to \infty$ のとき $1-\sqrt{2}/2 \approx 0.293$ に近づく．