統計検定　2級　解答解説（2018年6月実施：第12回）

問 1（解答番号 $\boxed{1}$〜$\boxed{2}$ ）

[1]

縦軸をみますと, III は総得点とわかります。

また, J1 の最高総得点 71 を標準化すると, $\dfrac{71-44.06}{12.89}=2.1$ となりますので, I が標準化得点になります。

$\boxed{1}$の正解は⑤となります。

[2]

$[54.45-2\times 11.77, 54.45+2\times 11.77]$, つまり $[30.91, 77.99]$ にない値がいくつあるか, ですが, 85 の 1 つだけです。

以上より, $\boxed{2}$の正解は②となります。

問 2（解答番号 $\boxed{3}$〜$\boxed{5}$ ）

[1]

I：東京都という外れ値を除いても, 正の相関があると言えます。正しいです。

II：正しいです。

III：「いかなる関係も」が誤っています。

$\boxed{3}$の正解は①となります。

[2]

I：北海道と同程度の人口の都道府県（600 万人くらい）では, 病床数は北海道より少ないです。誤りです。

II：人口と病床数は正の相関があるので, 人口が大きいほど病床数も多いです。よって, 病床数/人口 を考えると, その分散は病床の分散よりもずっと小さいと考えられます。

変動係数は 標準偏差/平均 です。分散がずっと小さくなっているとはいうものの, 病床数/人口では分母の平均も小さいため, 実際は計算して求めていかないといけないでしょうが, I と III が誤りであるため, 選択肢をみると, II が正しい, となります。

III：正の相関があります。誤りです。

$\boxed{4}$の正解は⑤となります。

[3]

I：偏相関係数の内容を正しく記述しています。正しいです。

II：正しいです。

III : 併設については何もわからないので, 誤りです。

$\boxed{5}$の正解は①となります。

問 3（解答番号 $\boxed{6}$〜$\boxed{8}$）

[1]

表の値から, ドイツの累積相対度数は,

$$8.4,\ 21.5,\ 38.7,\ 61.4,\ 100$$

となります。グラフの縦軸の値に一致します。なお, 値が非常に近いスウェーデンでは, 60 %のところで縦軸の値が 40 を越えます。

以上より, $\boxed{6}$の正解は⑤となります。

[2]

まず, 曲線下面積を求めます。なお, %表示はすべて小数にします。

$$\begin{aligned}
&0.2\times\frac{8.4}{100}\times\frac{1}{2}+0.2\times\left(\frac{8.4}{100}+\frac{21.5}{100}\right)\times\frac{1}{2}+0.2\times\left(\frac{21.5}{100}+\frac{38.7}{100}\right)\times\frac{1}{2}\\
&+0.2\times\left(\frac{38.7}{100}+\frac{61.4}{100}\right)\times\frac{1}{2}+0.2\times\left(\frac{61.4}{100}+\frac{100}{100}\right)\times\frac{1}{2}\\
&=0.36
\end{aligned}$$

よって, 完全平等線（対角線）とローレンツ曲線の間の面積は, $0.5-0.36=0.14$, ジニ係数はその 2 倍の 0.28 となります。

$\boxed{7}$の正解は②となります。

[3]

I : 正しいです。

II : アメリカのローレンツ曲線が, 完全平等線よりも最も遠くにあります。したがって, ジニ係数は最も大きいです。誤りです。なお, 最も不平等, という記述は合っています。

III : 正しいです。

以上より, $\boxed{8}$の正解は⑤となります。

問 4（解答番号 $\boxed{9}$〜$\boxed{10}$）

[1]

$$\frac{87.5-89.5}{89.5}\times 100\approx -2.2\ (\%)$$

となります。$\boxed{9}$の正解は③となります。

[2]

例えば, 2011 年のデータ x_{2011} は, r（%）を用いてどのように表されるか, ですが,

$$x_{2011} = 89.5\left(1 + \frac{r}{100}\right)$$

となります。同様に, 2015 年では,

$$x_{2015} = 89.5\left(1 + \frac{r}{100}\right)^5$$

となります。$x_{2015} = 100$ として, この式を, $r =$ の形に整理します。

$\boxed{10}$の正解は②となります。

問 5（解答番号 $\boxed{11}$ ）

フィッシャーの 3 原則は, 「繰り返し（反復）, 局所管理, 無作為化」です。

$\boxed{11}$の正解は②となります。

問 6（解答番号 $\boxed{12}$ ）

層化抽出法のうちの, 比例抽出法です。

$\boxed{12}$の正解は⑤となります。

問 7（解答番号 $\boxed{13}$〜$\boxed{14}$ ）

[1]

最初に T に勝ち, 次に U に勝つ (2 連勝したのでここで終わり）

もしくは,

最初に T に負け, 次に U に勝ち, 最後に T に勝つ

の 2 通りで, 互いに排反なので, 2 連勝する確率は,

$$pq + (1-p)qp$$

となります。$\boxed{13}$の正解は⑤となります。

[2]

同様に, $U - T - U$ の順に対戦して 2 連勝する確率は,

$$qp + (1-q)pq$$

です。ここで,

$$\{pq + (1-p)qp\} - \{qp + (1-q)pq\} = pq\{1 + (1-p) - 1 - (1-q)\} = pq(q-p) > 0$$

なので,

$$pq + (1-p)qp > qp + (1-q)pq$$

となります。$\boxed{14}$の正解は①となります。

問 8（解答番号 $\boxed{15}$〜$\boxed{17}$ ）

[1]

ある年 t における 6 月の電気料金を X_t としますと, X_t は, t 毎に独立に, 正規分布 $N(4000, 500^2)$ に従います。

$$P\left(X_t \geq 4800\right) = P\left(Z \geq \frac{4800-4000}{500}\right) = P(Z \geq 1.6) = 0.0548$$

となります（Z は標準正規分布に従う確率変数です。以下同じ)。

$\boxed{15}$の正解は②となります。

[2]

$X_t \sim N(4000, 500^2)$, $X_{t-1} \sim N(4000, 500^2)$ であり, 互いに独立ですので, $(Y = X_t - X_{t-1}) \sim N(0, 500^2 + 500^2)$ となります（$\sim$ は, 従う, の意です)。よって, 求める確率は,

$$P\left(Y \geq 800\right) = P\left(Z \geq \frac{800-0}{\sqrt{500^2+500^2}}\right) \approx P(Z \geq 1.13) = 0.1292$$

となります。$\boxed{16}$の正解は③となります。

[3]

問題文は,「正規分布 $N(4000, 500^2)$ から, 3 つの標本 X_t, X_{t-1}, X_{t-2} を無作為抽出ときの, X_t が最大である確率」を求めるものです。ところが, 取り出す標本が 3 つですと, X_t は最大値, 中央値, 最小値のいずれかになります。どれになるかは, 等確率なので, 求める答えは 1/3 になります。

$\boxed{17}$の正解は②となります。

問 9（解答番号 $\boxed{18}$〜$\boxed{19}$ ）

[1]

$$E[X^2] = V[X] + (E[X])^2 = 1.0 + 2.0^2 = 5.0$$
$$E[Y^2] = V[Y] + (E[Y])^2 = 1.0 + 3.0^2 = 10.0$$
$$Cov[X,Y] = E[XY] - E[X]E[Y] = 6.3 - 2.0 \times 3.0 = 0.3$$

となります。$\boxed{18}$の正解は④となります。

[2]

$$Cov[U,V] = Cov[3X-2, -2Y-4] = 3 \times (-2) \times Cov[X,Y] = -1.8$$

です。また, 相関係数は, もともとは,

$$r_{XY} = \frac{Cov[X,Y]}{\sqrt{V[X]}\sqrt{V[Y]}} = 0.3$$

です。値自体は線形変換により変わりませんが, 符号が変わります。

$\boxed{19}$の正解は④となります。

問 10（解答番号 $\boxed{20}$〜$\boxed{21}$）

[1]

X の標本平均 $\overline{X}$ は, 正規分布 $N\left(\mu, \dfrac{\sigma^2}{n}\right)$ に従います。いま, $\sigma^2 = 1$ です。よって, $Y = \overline{X} - \mu$ は正規分布 $N(0, 1/n)$ に従い,

$$\begin{aligned}
&P(|\overline{X} - \mu| \leq 0.5) \geq 0.95 \\
&P(-0.5 \leq Y \leq 0.5) \geq 0.95 \\
&P\left(\frac{-0.5}{\sqrt{1/n}} \leq Z \leq \frac{0.5}{\sqrt{1/n}}\right) \geq 0.95 \\
&P\left(-0.5\sqrt{n} \leq Z \leq 0.5\sqrt{n}\right) \geq 0.95
\end{aligned}$$

となります。この式を満たす最小の n は, $0.5\sqrt{n} = 1.96$（標準正規分布の上側 2.5 %点）を満たす n です。これを解き, $n = 15.3664$ を得ます。よって, 最小の整数 n は 16 です。

$\boxed{20}$の正解は④となります。

[2] μ の 95 %信頼区間（母分散未知の場合）は,

$$\overline{X} - t_{0.025}(n-1)\sqrt{\frac{U^2}{n}} < \mu < \overline{X} + t_{0.025}(n-1)\sqrt{\frac{U^2}{n}}$$

です。各値を代入します。

$\boxed{21}$の正解は①となります。

問 11（解答番号 $\boxed{22}$〜$\boxed{23}$）

[1]

母比率 p の 95 %信頼区間は, 標本サイズを n, 標本比率を $\hat{p}$ としたとき,

$$\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

で計算されます。$\hat{p} = 0.071$, $n = 4633$ を代入します。

$\boxed{22}$ の正解は③となります。

[2]

全体の母比率は, $\dfrac{N_1 p_1 + N_2 p_2}{N_1 + N_2}$ です。ここで, p_1, p_2 は未知なので, 標本比率 $\hat{p}_1$, $\hat{p}_2$ で代用して, 推定値とします。つまり, 推定値は, $\dfrac{N_1 \hat{p}_1 + N_2 \hat{p}_2}{N_1 + N_2}$ です。

標準誤差ですが, これは, 推定値の標準偏差のこと（ただし, その中に未知の母数があれば, その推定値で代用する）です。

$$V\left[\frac{N_1 \hat{p}_1 + N_2 \hat{p}_2}{N_1 + N_2}\right] = \left(\frac{N_1}{N_1 + N_2}\right)^2 V[\hat{p}_1] + \left(\frac{N_1}{N_1 + N_2}\right)^2 V[\hat{p}_2]$$

ですが, ここで, $V[\hat{p}_1] = \dfrac{p_1(1-p_1)}{n_1}$, $V[\hat{p}_2] = \dfrac{p_2(1-p_2)}{n_2}$ です。p_1, p_2 を $\hat{p}_1$, $\hat{p}_2$ で代用し, 平方根をとることで, 解答を得ます。

$\boxed{23}$ の正解は②となります。

問 12（解答番号 $\boxed{24}$〜$\boxed{25}$ ）

[1]

母平均の差の検定（対応がない場合）ですが, 帰無仮説（2 つの集団の母平均は等しい）のもとでの検定統計量は,

$$T = \frac{233.7 - 185.3}{\sqrt{\left(\dfrac{1}{6} + \dfrac{1}{6}\right)\dfrac{13549 + 7763}{6 + 6 - 2}}} \approx 1.82$$

となります。

$\boxed{24}$ の正解は④となります。

[2]

層内変動 S_W は,

$$S_W = 13549 + 7763 = 21312$$

です。また, 全平均 $\overline{X}$ が,

$$\overline{X} = \frac{6 \times 233.7 + 6 \times 185.3}{6 + 6} = 209.5$$

なので, 層間変動 S_B は,

$$S_B = 6(233.7 - 209.5)^2 + 6(185.3 - 209.5)^2 = 7027.68$$

となります。よって, 検定統計量である F 値は,

$$F = \frac{S_B/(k-1)}{S_W/(n-k)}$$

ですが, ここに $k=2$（層の数）, $n=12$（全数）, S_B, S_W を代入して, 約 3.30 となります。

$\boxed{25}$ の正解は④となります。

問 13（解答番号 $\boxed{26}$〜$\boxed{27}$）

[1]

第 1 種の過誤率は, 「H_0 が正しいときに H_0 を棄却する確率」なので, X が P_0 に従うときに $X \leq 3$ となる確率です。つまり, 0.3 です。

第 2 種の過誤率は, 「H_1 が正しいときに H_0 を棄却しない確率」なので, X が P_1 に従うときに $X > 3$ となる確率です。つまり, 0.1 です。

検出力は, $1 -$ 第 2 種の過誤率 $= 0.9$ です。

$\boxed{26}$ の正解は④となります。

[2]

同様に考えていきます。なお, 第 1 種の過誤率 $=$ 有意水準 です。

検定 II において,

第 1 種の過誤率 : 0.2, 第 2 種の過誤率 : 0.3, 検出力 : 0.7

検定 III において,

第 1 種の過誤率 : 0.3, 第 2 種の過誤率 : 1, 検出力 : 0

となります。なお, 検定 I は,

第 1 種の過誤率 : 0.3, 第 2 種の過誤率 : 0.1, 検出力 : 0.9

でした。

$\boxed{27}$ の正解は②となります。

問 14（解答番号 $\boxed{28}$〜$\boxed{30}$）

[1]

$$-7.08851 + 0.09408 \times 2.8 + 2.41815 \times 5.6 - 0.06498 \times 5.3 \approx 6.4$$

となります。$\boxed{28}$ の正解は⑤となります。

[2]

t 検定量は,

$$t = \frac{-0.06498 - (-0.5)}{0.22718} \approx 1.9$$

です。有意水準を α としたとき, 両側検定なので, 自由度 $43(= 47 - 3 - 1)$ の t 分布の上側 $\alpha/2$ 点と比較します。

帰無仮説が棄却されるには，この上側 $\alpha/2$ 点より t 値が大きくないといけません。t 分布表の自由度 40 辺りをみると，1.9 は上側 5 %点より大きいが，上側 2.5 %点よりは小さいことがわかります。よって，$\alpha/2 = 0.05$，つまり $\alpha = 0.1$ となります。

$\boxed{29}$ の正解は④となります。

[3]

I：P 値が 0.01 より小さい係数は，切片と log (賃金) の 2 つです。正しいです。

II：β_2 の推定値が正なので，

賃金が上がる →log(賃金) が大きくなる →log(犯罪発生率) が大きくなる → 犯罪発生率が増えるとなります（$\log x$ は単調増加関数です)。本文は誤りです。

III：「Adjusted R-squared」が，自由度修正済み決定係数を指します。正しいです。

$\boxed{30}$ の正解は④となります。

問 15（解答番号 $\boxed{31}$～$\boxed{33}$ ）

[1]

期待度数は，$\dfrac{(105+102)\times(105+15)}{365} \approx 68.05$ となります。

$\boxed{31}$ の正解は③となります。

[2]

独立性の χ^2 検定の検定統計量は，

$$\sum_{\text{セル}} \frac{(\text{観測度数} - \text{期待度数})^2}{\text{期待度数}}$$

で計算されます。

$\boxed{32}$ の正解は②となります。

[3]

独立性の χ^2 検定は，右片側検定です。また，帰無仮説は「関連がない」なので，それを棄却することは，「関連がある」と判断する，ということです。

$\boxed{33}$ の正解は⑤となります。

問 16（解答番号 $\boxed{34}$ ）

母分散の比の検定です。検定統計量は，各々の集団からの不偏分散 U_A^2, U_B^2 を用いて，

$$F = \frac{U_A^2}{U_B^2}$$

です。いま値を代入すると,

$$F = \frac{19.5^2}{14.5^2} \approx 1.81$$

となります。両側検定なので, 自由度 $(20, 40)$ の F 分布の上側 2.5 %点（2.068）と比較します。F 値はこの値より小さいので, 帰無仮説を棄却しません。

$\boxed{34}$の正解は②となります。

以上です。