

統計検定2級 全14回のオリジナル解説より抜粋

【第14回：2019年6月実施】

問3 (解答番号 12 ~ 13)

[1]

I: 正しいです。

z_i は標準化された得点なので、その平均は0になります。また、不偏分散の正の平方根を標準偏差として用い、標準化を行っていますので、 z_i の不偏分散の値は1です。平均が0なので、

$$\frac{1}{16} \sum_{i=1}^{17} z_i^2 = 1 \text{ になります。}$$

(補足)

数式で表現します。摂氏のデータを x_i 、不偏分散を U^2 と表しますと、

$$z_i = \frac{x_i - 2.4}{\sqrt{U^2}} = \frac{x_i - 2.4}{\sqrt{\frac{1}{16} \sum_{i=1}^{17} (x_i - 2.4)^2}}$$

なので、

$$\frac{1}{16} \sum_{i=1}^{17} z_i^2 = \frac{1}{16} \sum_{i=1}^{17} \left(\frac{(x_i - 2.4)^2}{\frac{1}{16} \sum_{i=1}^{17} (x_i - 2.4)^2} \right) = \frac{\frac{1}{16} \sum_{i=1}^{17} (x_i - 2.4)^2}{\frac{1}{16} \sum_{i=1}^{17} (x_i - 2.4)^2} = 1$$

となります。

II: 誤りです。

No.14 のデータについて、 $z_{14} = \frac{22 - 2.4}{7.0} = 2.8$ となり、2.5 を上回ります。

III: 正しいです。

摂氏のデータを x_i 、華氏のデータを y_i と表しますと、 $y_i = 1.8x_i + 32$ です。すると、 y_i の平均は $1.8 \times 2.4 + 32$ 、 y_i の不偏分散は x_i の 1.8^2 倍になるので、 y_i の標準偏差は 1.8×7.0 になります。よって、

$$w_i = \frac{y_i - (1.8 \times 2.4 + 32)}{1.8 \times 7.0} = \frac{(1.8x_i + 32) - (1.8 \times 2.4 + 32)}{1.8 \times 7.0} = \frac{x_i - 2.4}{7.0} = z_i$$

となります。7 の正解は④となります。

問 9 (解答番号 16 ~ 17)

[1]

まず, $Cov[X, Y] = E[XY] - E[X]E[Y] = 4 - 1 \times 2 = 2$ となります。また,

$$\begin{aligned} V[Z] &= V[X + Y] = V[X] + V[Y] + 2Cov[X, Y] = V[X] + V[Y] + 4 \\ V[W] &= V[2X - Y] = V[2X] + V[-Y] + 2Cov[2X, -Y] \\ &= 4V[X] + V[Y] - 4Cov[X, Y] = 4V[X] + V[Y] - 8 \end{aligned}$$

より, $V[Z] = V[W] = 24$ を $V[X], V[Y]$ の連立方程式として解きますと, $V[X] = 4, V[Y] = 16$ となります。よって, $V[X] = E[X^2] - (E[X])^2, V[Y] = E[Y^2] - (E[Y])^2$ から, $E[X^2] = 5, E[Y^2] = 20$ となります。

16 の正解は ③ となります。

[2]

$$\frac{Cov[X, Y]}{\sqrt{V[X]}\sqrt{V[Y]}} = \frac{2}{2 \times 4} = 0.25$$

となります。17 の正解は ④ となります。

問 16 (解答番号 28 ~ 30)

[1]

第 1 種の過誤率を求めます。 H_0 が正しいとしたとき, つまり $X \sim N(0, 1)$ のとき, H_0 を棄却する確率なので, 標準正規分布表を用いて, その確率は, $P(X \geq 0.8) = 0.2119$ となります。

第 2 種の過誤率については, H_1 が正しいとしたとき, つまり $X \sim N(1, 1)$ のとき, H_0 を採択する確率なので,

$$\begin{aligned} P(X < 0.8) &= P\left(\frac{X - 1}{\sqrt{1}} < \frac{0.8 - 1}{\sqrt{1}}\right) \\ &= P(Z < -0.2) \quad (Z \text{ は標準正規分布に従う確率変数}) \\ &= 0.4207 \end{aligned}$$

となります。

28 の正解は ② となります。

[2]

[1] を参考にしますと,

$$\begin{aligned} x_0 \text{ 増加} &\rightarrow \alpha(x_0) \text{ 減少} \rightarrow 1 - \alpha(x_0) \text{ 増加} \\ x_0 \text{ 増加} &\rightarrow \beta(x_0) \text{ 増加} \end{aligned}$$

となりますので、 $1 - \alpha(x_0)$ が大きくなれば、 $\beta(x_0)$ も大きくなります。よって、①か②になります。ここで、直線関係ではないであろう、と推測して、①を選ぶことができます。

さらに、標準正規分布の確率密度関数を $\phi(x)$ 、分布関数を $\Phi(x)$ と表しますと、

$$\alpha(x_0) = 1 - \Phi(x_0), \quad \beta(x_0) = \Phi(x_0 - 1)$$

となります。そこで、グラフの縦軸と横軸について、 $y = 1 - \alpha(x_0) = \Phi(x_0)$ 、 $x = \beta(x_0) = \Phi(x_0 - 1)$ としますと、 $y/x = \Phi(x_0)/\Phi(x_0 - 1)$ は明らかに定数ではないです。したがって、直線関係ではなく、①を選ぶことができます。

(補足：グラフが上に凸であることについて)

$$\frac{dy}{dx_0} = \phi(x_0), \quad \frac{dx}{dx_0} = \phi(x_0 - 1) \text{ なので,}$$

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy/dx_0}{dx/dx_0} = \frac{\phi(x_0)}{\phi(x_0 - 1)} = \frac{e^{-x_0^2/2}}{e^{-(x_0-1)^2/2}} = e^{-\frac{1}{2}(2x_0-1)} = e^{\frac{1}{2}e^{-x_0}} > 0 \\ \frac{d^2y}{dx^2} &= \frac{d}{dx} \left(\frac{dy}{dx} \right) = \frac{dx_0}{dx} \frac{d}{dx_0} \left(\frac{dy}{dx} \right) = \frac{e^{\frac{1}{2}(-e^{-x_0})}}{\phi(x_0 - 1)} < 0 \end{aligned}$$

より、上に凸になります。

29の正解は①となります。

[3]

$\alpha(x_0) + \beta(x_0) = 1 - \Phi(x_0) + \Phi(x_0 - 1)$ であり、この値は、 x_0 が増加するにつれ、 $\Phi(x_0)$ 、 $\Phi(x_0 - 1)$ ともに増加することから、最小値が 0 から 1 の間にある、と推測されます。

(補足)

$\alpha(x_0) + \beta(x_0) = 1 - y + x$ ですが、ここで $1 - y + x = k$ 、すなわち $y = x + 1 - k$ とおきます。 k が最小値をとるとき、 $1 - k$ は最大値をとります。ここで、①のグラフと、直線 $y = x + 1 - k$ が接するとき、その切片 $1 - k$ は最大になり、 k は最小になります。

接するときの接線の傾きについて、 $e^{\frac{1}{2}e^{-x_0}} = 1$ なので、 $x_0 = 1/2$ となります。

30の正解は②となります。

問 17 (解答番号 31 ~ 33)

[1]

I: 誤りです。 $C = U = G = 0$ とすれば、 $y = \beta_1 + u$ となりますが、このときの y が高校卒の学歴における初任給となります。

II: 正しいです。大学卒 ($C = 0, U = 1, G = 0$) では $y = \beta_1 + \beta_3 + u$ 、大学院修士課程修了 ($C = 0, U = 0, G = 1$) では $y = \beta_1 + \beta_4 + u$ となります。推定量を代入して差をみます。

III : 誤りです。 t 分布の自由度は、 $n - k - 1$ です (n は標本数、 k は説明変数の数で、 いまは 3 です)。 よって、 $16 - 3 - 1 = 12$ より、 自由度は 12 になります。

31 の正解は ② となります。

[2]

I : 正しいです。

II : 誤りです。 回帰残差の 2 乗和を Q としますと、 決定係数は

$$1 - \frac{Q}{\sum_i (y_i - \bar{y})^2}$$

となり、 自由度修正済み決定係数は、 説明変数の個数を k として、

$$1 - \frac{\frac{1}{n - k - 1} Q}{\frac{1}{n - 1} \sum_i (y_i - \bar{y})^2}$$

となります。 単回帰モデルでは $k = 1$ なので、 これら 2 つの値は異なります。

III : 誤りです。 両側検定においては、 両側 P 値は、 $t > 11.109$ である確率の 2 倍にします。 一方で片側検定においては、 $t > 11.109$ である確率が P 値になります。

32 の正解は ① となります。

[3]

I : 誤りです。 説明変数が増えるにつれ、 決定係数は小さくなることはありません。 モデルの選択法の一つとしては、 自由度修正済み決定係数を用います。

II : 正しいです。 学歴ダミー変数の方は、 回帰係数がそれぞれ異なります。 一方、 教育年数の方は、 x が 2 ずつ増えるので、 初任給は 1.187×2 (万円) ずつ増えます。

III : 正しいです。

33 の正解は ⑤ となります。

【第 13 回 : 2018 年 11 月実施】

問 8 (解答番号 12 ~ 13)

[1]

Y は平均 $0.3 + 2x$, 分散 1 の正規分布に従います。

$$P(Y \geq 0) = P(0.3 + 2x + U \geq 0) = P(U \geq -2x - 0.3)$$

この確率が 0.95 です。標準正規分布の上側 95 % 点が -1.645 なので, $-2x - 0.3 = -1.645$ を解いて, $x \approx 0.67$ となります。12 の正解は④となります。

[2]

Y の上側 5 % 点を y と表しますと, $P(Y \geq y) = P(0.3 + 2x + U \geq y) = P(U \geq y - 2x - 0.3)$ となり, この確率が 0.05 なので, 今度は標準正規分布の上側 5 % 点 1.645 を用いて, $y - 2x - 0.3 = 1.645$, つまり $y = 2x + 1.945$ となります。 x と y が直線関係にあるグラフを選びます。

13 の正解は①となります。

問 9 (解答番号 14 ~ 15)

[1]

$$\frac{P(X = x + 1)}{P(X = x)} = \frac{{}_7C_{x+1}(2/3)^{6-x}(1/3)^{x+1}}{{}_7C_x(2/3)^{7-x}(1/3)^x} = \frac{\frac{7!}{(x+1)!(6-x)!} \cdot \frac{1}{3}}{\frac{7!}{x!(7-x)!} \cdot \frac{2}{3}} = \frac{-x+7}{2(x+1)}$$

より, $a = 7, b = 2$ となります。14 の正解は②となります。

[2]

$\frac{P(X = x + 1)}{P(X = x)} > 1$ を解くと $3x < 5$ となりますので, $x = 0, 1$ のとき, $P(X = x + 1) > P(X = x)$ です。つまり, $P(X = 0) < P(X = 1) < P(X = 2)$ です。

同様に, $\frac{P(X = x + 1)}{P(X = x)} < 1$ より $x \geq 2$ となります。 $x \geq 2$ のとき, $P(X = x + 1) < P(X = x)$ です。つまり, $P(X = 2) > P(X = 3) > \dots$ となります。

以上より, $x = 2$ で $P(X = x)$ は最大となります。15 の正解は②となります。

問 11 (解答番号 17~19)

[1]

正規分布の歪度は 0, 尖度は, 問題文の定義によりますと, 0 です (正規分布の μ_4/σ^4 の値は常に 3 です)。17 の正解は①となります。

[2]

一様分布は平均に関して対称ですので, 歪度は 0 になります。尖度は, 正規分布よりも中心部が平坦で, 裾が短いため, 負の値をとると考えられます。念のため計算しますと, 平均が 0 であることから,

$$\sigma^2 = E[X^2] = \int_{-1}^1 x^2 \cdot \frac{1}{2} dx = \frac{1}{3}$$
$$E[(X - \mu)^4] = E[X^4] = \int_{-1}^1 x^4 \cdot \frac{1}{2} dx = \frac{1}{5}$$

より, 尖度は $(1/5)/(1/3^2) - 3 = -1.2$ となります。18 の正解は⑤となります。

[3]

I: 逆です。右に裾が長いと, 平均より右側の (平均より大きい) データの影響が大きくなり, $(X - \mu)^3$ の値は正に大きくなっていく, と考えるとよいです。

II: 逆です。裾が長いほど, 絶対値の大きいデータの影響を受け, 大きくなります。

III: t 分布は平均に関して対称なので, 歪度は 0, ここまでは正しいです。自由度が大きくなるほど, 標準正規分布 (尖度 0) に近づいていきますので, 尖度の絶対値は小さくなります。

19 の正解は⑤となります。

問 14 (解答番号 22~23)

[1]

母分散の比の検定です。第 1 自由度は $30 - 1$, 第 2 自由度は $31 - 1$ です。22 の正解は⑤となります。

[2]

第 1 種の過誤の確率とは, 帰無仮説 H_0 が正しいのに, H_0 を棄却する確率です。その確率ですが, H_0 が正しいとき,

$$\begin{aligned} & \text{(3つの検定のどれか1つでも } H_0 \text{ を棄却する確率)} \\ &= 1 - \text{(3つの検定のいずれも } H_0 \text{ を棄却しない確率)} \\ &= 1 - \text{(1つの組み合わせの検定で } H_0 \text{ を棄却しない確率)}^3 \\ &= 1 - (1 - 0.05)^3 \approx 0.14 \end{aligned}$$

となります。23 の正解は④となります。

問 15 (解答番号 24 ~ 26)

[1]

X は二項分布 $B(200, 0.05)$ に従います。その期待値は $200 \times 0.05 = 10$ 、分散は $200 \times 0.05 \times (1 - 0.05) = 9.5$ となります。24 の正解は③となります。

[2]

不良品の母比率を r 、標本比率を \hat{r} とすると、 \hat{r} は、標本数 n が十分大きいとき、平均 r 、分散 $r(1 - r)/n$ の正規分布に従うと近似できます。帰無仮説 H_0 のもとで、 $r = 0.05$ 、 $n = 200$ を代入しますと、その平均は 0.05 、分散は 0.0002375 となります。

さて、対立仮説より右片側検定を行いますので、片側 P 値は、

$$P(\hat{r} > 16/200) = P\left(\frac{\hat{r} - 0.05}{\sqrt{0.0002375}} > \frac{16/200 - 0.05}{\sqrt{0.0002375}}\right) \approx P(Z > 1.95) \approx 0.26$$

となります (Z は標準正規分布に従う確率変数)。25 の正解は②となります。

[3]

それぞれのメーカーの標本比率は、 $\hat{r}_A = 16/200$ 、 $\hat{r}_B = 17/200$ です。よって、 $\hat{d} = \hat{r}_A - \hat{r}_B = -0.005$ となります。

ところで、 \hat{r}_A は正規分布 $N\left(r_A, \frac{r_A(1 - r_A)}{n_A}\right)$ に、 \hat{r}_B は正規分布 $N\left(r_B, \frac{r_B(1 - r_B)}{n_B}\right)$ に従うので、 \hat{d} は正規分布 $N\left(r_A - r_B, \frac{r_A(1 - r_A)}{n_A} + \frac{r_B(1 - r_B)}{n_B}\right)$ に従います (なお、 $n_A = n_B = 200$)。

帰無仮説 H_0 : 「 $d = r_A - r_B = 0$ 」のもとで、分散の中で $r_A = r_B = \frac{16 + 17}{200 + 200} = 0.0825$ と近似すると、 \hat{d} は、正規分布 $N(0, 0.000825)$ に従います。

対立仮説より両側検定なので、求める両側 P 値は、

$$2 \times P(\hat{d} < -0.005) = 2 \times P\left(Z < -\frac{0.005}{\sqrt{0.000825}}\right) \approx 2 \times P(Z < -0.174) \approx 2 \times 0.43 = 0.86$$

と求められます。26 の正解は⑤となります。

問 18 (解答番号 32 ~ 34)

[1]

I : 正しいです。残差平方和を S としますと、

$$\hat{\sigma}_u^2 = \frac{S}{n - k - 1}$$

という関係があります。 n は標本数でいまは 5 、 k は説明変数の数でいまは 1 です。よって、 $S = (5 - 1 - 1) \times 0.608^2 \approx 1.1$ となります。

II : 誤りです。 t 値は変わりません。

III : 正しいです。

32の正解は④となります。

[2]

I：誤りです。 z の P 値は、切片や x の P 値と比べて小さいです。

II：誤りです。説明変数間の相関が高いと、多重共線性の問題が生じます。

III：誤りです。 P 値の方が有意水準より大きいため、帰無仮説を棄却しません。

33の正解は⑤となります。

[3]

I：誤りです。同じ値になるとは限りません。

II：正しいです。

III：誤りです。 b' が有意でないので、「 x が変化しても y は変化しない」と解釈されます。

34の正解は①となります。

【第12回：2018年6月実施】

問7 (解答番号 13～14)

[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。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。14の正解は①となります。

問8 (解答番号 15～17)

[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。

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

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)。

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、

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

となります。[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} を無作為抽出するときの、 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。[17]の正解は②となります。

問 9 (解答番号 [18] ~ [19])

[1]

$$E[X^2] = V[X] + (E[X])^2 = 1.0 + 2.0^2 = 5.0$$

$$E[Y^2] = V[Y] + (E[Y])^2 = 1.0 + 3.0^2 = 10.0$$

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 6.3 - 2.0 \times 3.0 = 0.3$$

となります。[18]の正解は④となります。

[2]

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

です。また、相関係数は、もともとは、

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

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。

[19]の正解は④となります。

問 11 (解答番号 [22] ~ [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。

22の正解は③となります。

[2]

全体の母比率は、 $\frac{N_1 p_1 + N_2 p_2}{N_1 + N_2}$ 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、 $\frac{N_1 \hat{p}_1 + N_2 \hat{p}_2}{N_1 + N_2}$ 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。

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

ですが、ここで、 $V[\hat{p}_1] = \frac{p_1(1-p_1)}{n_1}$, $V[\hat{p}_2] = \frac{p_2(1-p_2)}{n_2}$ 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。

23の正解は②となります。

問 13 (解答番号 26 ~ 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、 $1 - \text{第 2 種の過誤率} = 0.9$ です。

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。

27の正解は②となります。

【第 11 回 : 2017 年 11 月実施】

問 8 (解答番号 15 ~ 16)

[1]

全確率が 1 になるようにします。

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^2 cx(2-x)dx = c \left[x^2 - \frac{1}{3}x^3 \right]_0^2 = \frac{4}{3}c$$

で, $\frac{4}{3}c = 1$ より, $c = \frac{3}{4}$ となります。

15 の正解は③となります。

[2] 平均 $E[X]$ は,

$$E[X] = \int_{-\infty}^{\infty} x \times f(x)dx = \frac{3}{4} \int_0^2 x^2(2-x)dx = \frac{3}{4} \left[\frac{2}{3}x^3 - \frac{1}{4}x^4 \right]_0^2 = 1$$

となります。次に,

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \times f(x)dx = \frac{3}{4} \int_0^2 x^3(2-x)dx = \frac{3}{4} \left[\frac{1}{2}x^4 - \frac{1}{5}x^5 \right]_0^2 = \frac{6}{5}$$

ですので,

$$V[X] = E[X^2] - (E[X])^2 = \frac{6}{5} - 1^2 = \frac{1}{5}$$

となります。

16 の正解は④となります。

問 9 (解答番号 17 ~ 18)

[1]

正規分布から派生する分布の定義の通りです。 17 の正解は⑤となります。

[2]

Y は第一自由度 20, 第二自由度 10 の F 分布に従います。その下側 5% 点は,

$$\frac{1}{F_{0.05}(10, 20)} = \frac{1}{2.348}$$

となります。

18 の正解は②となります。

問 10 (解答番号 19 ~ 21)

[1]

$$P(X_1 \geq 60) = P\left(Z \geq \frac{60 - 50}{10}\right) = P(Z \geq 1) = 0.1587$$

となります (Z は標準正規分布に従う確率変数)。

19 の正解は ① となります。

[2]

どの 1 人が 60 点以上かで 5 通りありますので、 $X \sim N(50, 10^2)$ として、

$$5 \times P(X \geq 60) \times \{P(X < 60)\}^4 = 5 \times 0.1587 \times (1 - 0.1587)^4 \approx 0.40$$

となります。

20 の正解は ④ となります。

[3]

5 人の標本平均 $\bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$ は、 $N(50, 10^2/5)$ に従います。よって、

$$P(\bar{X} \geq 52) = P\left(Z \geq \frac{52 - 50}{\sqrt{10^2/5}}\right) = P(Z \geq 1/\sqrt{5}) \approx 0.33$$

となります。

21 の正解は ② となります。

問 13 (解答番号 27 ~ 28)

[1]

母比率の 95 % 信頼区間は、標本比率を \hat{p} として、

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

となります。 $n = 1897$, $\hat{p} = 0.483$ を代入します。

27 の正解は ③ となります。

[2]

ア：まず信頼区間ですが、ここでは、以下のように算出しています。平成 25 年における標本数を $n_A (= 1897)$ 、母比率を p_A 、標本比率を $\hat{p}_A (= 0.483)$ 、平成 21 年における標本数を $n_B (= 1925)$ 、母比率を p_B 、標本比率を $\hat{p}_B (= 0.416)$ とすると、

$$\left[(\hat{p}_A - \hat{p}_B) - 1.96 \times \sqrt{\frac{\hat{p}_A(1 - \hat{p}_A)}{n_A} + \frac{\hat{p}_B(1 - \hat{p}_B)}{n_B}}, (\hat{p}_A - \hat{p}_B) + 1.96 \times \sqrt{\frac{\hat{p}_A(1 - \hat{p}_A)}{n_A} + \frac{\hat{p}_B(1 - \hat{p}_B)}{n_B}} \right]$$

となります。各値を代入しますと、①か②に絞られます。

イ：アの信頼区間を計算すると、およそ、 $[0.036, 0.098]$ となり、0 を含まないことから、変化したといえます。

なお、以下のように母比率の差の検定を行っても、帰無仮説 H_0 「変化していない」を棄却できません。

「帰無仮説 H_0 「変化していない」のもとでの検定統計量は、

$$Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right) \hat{p}(1 - \hat{p})}}$$

であり、この Z は標準正規分布に従うとしてよい。なお、 \hat{p} は、 H_0 のもとでの標本比率であり、 $\hat{p} = \frac{n_A \times \hat{p}_A + n_B \times \hat{p}_B}{n_A + n_B}$ である。これらの値を代入すると、 $Z \approx 4.2$ である。両側検定を行い、標準正規分布の上側 2.5 % 点 1.96 より Z の方が大きいので、 H_0 を棄却する。」

28 の正解は②となります。

問 15 (解答番号 30 ~ 31)

[1]

一等が出る人数の期待度数 E は、 $50 \times 0.20 = 10$ 、二等、外れではそれぞれ $50 \times 0.30 = 15$ 、 $50 - 10 - 15 = 25$ です。まとめると、次の表のようになります (O は観測度数)。

	一等	二等	外れ
E	10	15	25
O	5	12	33

適合度検定の検定統計量は、

$$\chi^2 = \frac{(10 - 5)^2}{10} + \frac{(15 - 12)^2}{15} + \frac{(25 - 33)^2}{25} = 2.5 + 0.6 + 2.56 = 5.66$$

であり、用いる χ^2 分布の自由度は $(3 - 1) = 2$ です。自由度 2 の χ^2 分布の上側 5 % 点 5.99 よりこの検定統計量の値は小さいので、帰無仮説 H_0 を採択します。

30 の正解は②となります。

[2]

以上より、31 の正解は⑤となります。

【第10回：2017年6月実施】

問6 (解答番号 13)

$$X = a + b + \varepsilon_1, \quad Y = a - b + \varepsilon_2$$

より,

$$\frac{X+Y}{2} = a + \frac{\varepsilon_1 + \varepsilon_2}{2}, \quad \frac{X-Y}{2} = b + \frac{\varepsilon_1 - \varepsilon_2}{2}$$

ですので,

$$\begin{aligned} V\left[\frac{X-Y}{2}\right] &= V\left[b + \frac{\varepsilon_1 - \varepsilon_2}{2}\right] = V\left[\frac{\varepsilon_1 - \varepsilon_2}{2}\right] \\ &= \frac{1}{4}(V[\varepsilon_1] + V[\varepsilon_2]) \quad (\varepsilon_1 \text{ と } \varepsilon_2 \text{ は独立なので}) \\ &= \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2} \end{aligned}$$

となります。13の正解は③となります。

問9 (解答番号 19 ~ 20)

[1]

$$E[U] = E[X + Y] = E[X] + E[Y] = 0 + 0 = 0$$

$$E[V] = E[X - Y] = E[X] - E[Y] = 0 - 0 = 0$$

です。また、 X と Y は独立なので,

$$V[U] = V[X + Y] = V[X] + V[Y] = \sigma_1^2 + \sigma_2^2$$

$$V[V] = V[X - Y] = V[X] + V[Y] = \sigma_1^2 + \sigma_2^2$$

です。また,

$$\begin{aligned} Cov(U, V) &= Cov(X + Y, X - Y) = Cov(X + Y, X) - Cov(X + Y, Y) \\ &= Cov(X, X) + Cov(Y, X) - Cov(X, Y) - Cov(Y, Y) \\ &= V[X] - V[Y] = \sigma_1^2 - \sigma_2^2 \end{aligned}$$

です。よって、相関係数は,

$$r = \frac{Cov(U, V)}{\sqrt{V[U]}\sqrt{V[V]}} = \frac{\sigma_1^2 - \sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

となります。19の正解は④となります。

[2]

I: 正しいです。どちらも0です。

II: [1]の結果より、 $\sigma_1^2 = \sigma_2^2$ のとき、 $r = 0$ となります。 U, V はともに正規分布に従いますが、その場合は、 $r = 0$ ならば U と V は独立です（無相関ならば独立、というのは、いつも成り立つとは限りませんので、ご注意ください）。正しいです。

III: とともに、期待値0、分散 $\sigma_1^2 + \sigma_2^2$ の正規分布に従います。正しいです。

20の正解は⑤となります。

問10（解答番号 21～22）

[1]

W_1 は自由度1の χ^2 分布に従います。その上側5%点は3.84となります。

21の正解は④となります。

[2]

W_n は自由度 n の χ^2 分布に従います。

$n = 3$ のとき、自由度3の χ^2 分布の上側5%点7.81より、 $W_3 \geq 6$ となる確率は、5%より大きいです。

$n = 8$ のとき、自由度8の χ^2 分布の上側5%点15.51より、 $W_8 \geq 16$ となる確率は、5%より小さいです。

22の正解は②となります。

問11（解答番号 23）

母平均を μ 、標本平均（母平均の推定値）を \bar{X} 、母分散を σ^2 、標本数を n とします。

まず、母変動係数が0.4なので、 $\frac{\sigma}{\mu} = 0.4$ 、つまり $\mu = \sigma/0.4$ です。また、問題文の内容を式に表しますと、

$$Pr\left(-0.05 \leq \frac{\bar{X} - \mu}{\mu} \leq 0.05\right) = 0.95$$

となります。この分母の μ を $\sigma/0.4$ に置き換え、整理しますと、

$$Pr\left(-\frac{1}{8} \leq \frac{\bar{X} - \mu}{\sigma} \leq \frac{1}{8}\right) = 0.95$$

となります。ところで \bar{X} は期待値 μ 、分散 σ^2/n の正規分布に従いますので、括弧内の不等式の辺々

に \sqrt{n} を掛けて、

$$Pr\left(-\frac{\sqrt{n}}{8} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \underset{N(0,1) \text{ に従う}}{}} \leq \frac{\sqrt{n}}{8}\right) = 0.95$$

としますと、 $\frac{\sqrt{n}}{8}$ は標準正規分布の上側 2.5 %点 1.96 に等しいことがわかります。よって $n \approx 250$ となります。

23 の正解は④となります。

【第9回：2016年11月実施】

問7 (解答番号 12 ~ 13)

[1]

$$\begin{aligned}
 & \underbrace{\frac{4}{10}}_{\text{通りかかった人が A}} \times \underbrace{\frac{40}{100}}_{\text{A 型と予想する}} + \underbrace{\frac{3}{10}}_{\text{通りかかった人が B}} \times \underbrace{\frac{30}{100}}_{\text{B 型と予想する}} + \underbrace{\frac{2}{10}}_{\text{通りかかった人が O}} \times \underbrace{\frac{20}{100}}_{\text{O 型と予想する}} \\
 & + \underbrace{\frac{1}{10}}_{\text{通りかかった人が AB}} \times \underbrace{\frac{10}{100}}_{\text{AB 型と予想する}} \\
 & = \frac{3}{10}
 \end{aligned}$$

となります。12の正解は②となります。

[2]

$$\begin{aligned}
 & \frac{P(\text{実際に A 型である} \mid \text{A 型であると予想する})}{P(\text{実際に A 型である, かつ, A 型であると予想する})} \\
 & = \frac{P(A \text{ 型であると予想する})}{P(A \text{ 型であると予想する})} \\
 & = \frac{4/10 \times 2/3}{\underbrace{4/10}_{\text{A 型である}} \times \underbrace{2/3}_{\text{A 型の人を A 型と予想}} + \underbrace{6/10}_{\text{A 型でない}} \times \underbrace{1/3}_{\text{A 型でない人を A 型と予想}}} \\
 & = \frac{4}{7} \approx 0.57
 \end{aligned}$$

となります。13の正解は③となります。

問8 (解答番号 14 ~ 15)

[1]

$$\begin{aligned}
 \text{Cov}(X_1, Y) &= \text{Cov}\left(X_1, \frac{X_1 + X_2 + X_3}{3}\right) \\
 &= \frac{1}{3} \text{Cov}(X_1, X_1 + X_2 + X_3) \\
 &= \frac{1}{3} \{ \underbrace{\text{Cov}(X_1, X_1)}_{=V[X_1]=1} + \underbrace{\text{Cov}(X_1, X_2)}_{=0} + \underbrace{\text{Cov}(X_1, X_3)}_{=0} \} \\
 &= \frac{1}{3}
 \end{aligned}$$

であり, $V[X_1] = 1$,

$$V[Y] = V\left[\frac{X_1 + X_2 + X_3}{3}\right] = \frac{1}{9}V[X_1 + X_2 + X_3] = \frac{1}{9}(V[X_1] + V[X_2] + V[X_3]) = \frac{1}{3}$$

なので, 相関係数は,

$$\frac{Cov(X_1, Y)}{\sqrt{V[X_1]}\sqrt{V[Y]}} = \frac{1/3}{\sqrt{1/3}} = \frac{\sqrt{3}}{3} \approx 0.6$$

となります。[14]の正解は④となります。

[2]

$$\begin{aligned} Cov(X_1, Y) &= Cov\left(X_1, \frac{X_1 + X_2 + X_3}{3}\right) \\ &= \frac{1}{3}Cov(X_1, X_1 + X_2 + X_3) \\ &= \frac{1}{3}\{ \underbrace{Cov(X_1, X_1)}_{=V[X_1]=1} + \underbrace{Cov(X_1, X_2)}_{=0.5} + \underbrace{Cov(X_1, X_3)}_{=0.5} \} \\ &= \frac{2}{3} \end{aligned}$$

と,

$$\begin{aligned} V[Y] &= \frac{1}{9}V[X_1 + X_2 + X_3] \\ &= \frac{1}{9}\{V[X_1] + V[X_2] + V[X_3] + 2Cov(X_1, X_2) + 2Cov(X_2, X_3) + 2Cov(X_3, X_1)\} \\ &= \frac{2}{3} \end{aligned}$$

より, 相関係数は,

$$\frac{Cov(X_1, Y)}{\sqrt{V[X_1]}\sqrt{V[Y]}} = \frac{2/3}{\sqrt{2/3}} = \sqrt{\frac{2}{3}} \approx 0.8$$

となります。[15]の正解は⑤となります。

問9 (解答番号 [16] ~ [18])

[1]

ポアソン分布では, 母数 = 平均 = 分散 です。

[16]の正解は①となります。

[2]

イベントの参加者数を X と表しますと、 X の平均と分散はともに 50 であり、また、中心極限定理により、 X は正規分布に従うと近似できます。よって、

$$P(X > 60) = P\left(\frac{X - 50}{\sqrt{50}} > \frac{60 - 50}{\sqrt{50}}\right) \approx P(Z > 1.41) \approx 0.07$$

となります (Z は標準正規分布に従う確率変数)。17 の正解は⑤となります。

[3]

事前登録をしない参加者数を Y 、参加者数全体の人数を W としますと、 $W = Y + 30$ となります。2 と同様に、 Y は平均と分散が 20 の正規分布に従うと近似しますと、 W は平均 50、分散 20 の正規分布に従います。よって、

$$\begin{aligned} P(W > x) &= 0.05 \\ \therefore P\left(\frac{W - 50}{\sqrt{20}} > \frac{x - 50}{\sqrt{20}}\right) &= 0.05 \\ \therefore P\left(Z > \frac{x - 50}{\sqrt{20}}\right) &= 0.05 \quad (Z \text{ は標準正規分布に従う確率変数}) \\ \therefore \frac{x - 50}{\sqrt{20}} &= 1.645 \quad (\text{標準正規分布の上側 } 5\% \text{ 点}) \\ \therefore x &\approx 57 \end{aligned}$$

となります。18 の正解は③となります。

問 10 (解答番号 19 ~ 20)

[1]

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_{-1}^0 x^2(x+1) dx + \int_0^1 x^2(-x+1) dx \\ &= \left[\frac{1}{4}x^4 + \frac{1}{3}x^3\right]_{-1}^0 + \left[-\frac{1}{4}x^4 + \frac{1}{3}x^3\right]_0^1 \\ &= \frac{1}{6} \end{aligned}$$

となります。 $E[X] = 0$ より、 $V[X] = E[X^2] - (E[X])^2 = \frac{1}{6}$ となります。

19 の正解は①となります。

[2] a を第 1 四分位数としますと、 $\int_{-\infty}^a f(x) dx = \frac{1}{4}$ を解くこととなります。明らかに a は

$-1 \leq x < 0$ の範囲にありますので、

$$\int_{-\infty}^a f(x)dx = \int_{-1}^a (x+1)dx = \left[\frac{1}{2}x^2 + x \right]_{-1}^a = \frac{1}{2}a^2 + a + \frac{1}{2}$$

より、 $\frac{1}{2}a^2 + a + \frac{1}{2} = \frac{1}{4}$ つまり $2a^2 + 4a + 1 = 0$ を解いて、 $-1 \leq a < 0$ より、 $a = \frac{-2 + \sqrt{2}}{2} = -1 + \frac{\sqrt{2}}{2}$ となります。

20の正解は③となります。

問 11 (解答番号 21 ~ 22)

[1]

分散に関する式 $V[X_i] = E[X_i^2] - (E[X_i])^2$ と、 $E[X_i] = \mu$ 、 $V[X_i] = \sigma^2$ から、

$$E[X_i^2] = (E[X_i])^2 + V[X_i] = \mu^2 + \sigma^2$$

となります。21の正解は③となります。

[2]

(ア) には、不偏分散が入ります。①か②になります。

(イ) に関して、 $V[\bar{X}] = E[\bar{X}^2] - (E[\bar{X}])^2$ と、 $E[\bar{X}] = \mu$ 、 $V[\bar{X}] = \sigma^2/n$ から、

$$E[\bar{X}^2] = (E[\bar{X}])^2 + V[\bar{X}] = \mu^2 + \frac{\sigma^2}{n}$$

ですので、

$$E \left[\bar{X}^2 - \frac{\hat{\sigma}^2}{n} \right] = E[\bar{X}^2] - \frac{1}{n}E[\hat{\sigma}^2] = \mu^2 + \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = \mu^2$$

となります。22の正解は②となります。

問 13 (解答番号 25 ~ 27)

[1]

観測度数を表にまとめると、以下になります。

	60~100	100~120	120~140	140~160	160~200	合計
夏季	18	42	24	11	1	96
冬季	13	25	23	19	16	96
合計	31	67	47	30	17	192

となります。求める期待度数は、 $67 \times 96/192 = 33.5$ となります。

25の正解は④となります。

[2]

上記の分割表を用いた、独立性の χ^2 検定です。自由度は $(2 - 1) \times (5 - 1) = 4$ となります。

26の正解は②となります。

[3]

自由度 4 の χ^2 分布の上側 5 %点 9.49 と比較して、20.51 の方が大きいため、帰無仮説「2つの分布は同等である」を棄却します。

27の正解は①となります。

【第 8 回 : 2016 年 6 月実施】

問 3 (解答番号 6 ~ 8)

[1]

問題文を表に表しますと、次のようになります。

人数	はい	いいえ	合計
属性 A			40
属性 B			60
合計	20	80	100

独立性の仮定のもとでは、属性 B, いいえの欄の期待度数は $\frac{80 \times 60}{100} = 48$ となります。

6 の正解は ④ となります。

[2]

$(x, y) = (\text{属性 } A/B, \text{はい/いいえ})$ としますと、 $(0, 0)$ が 40 個, $(0, 1)$ が 10 個, $(1, 0)$ が 10 個, $(1, 1)$ が 40 個となります。 x の平均を \bar{x} , などと表しますと,

$$\begin{aligned}\bar{x} &= \frac{0 \times 40 + 0 \times 10 + 1 \times 10 + 1 \times 40}{100} = \frac{1}{2} \\ \bar{y} &= \frac{0 \times 40 + 1 \times 10 + 0 \times 10 + 1 \times 40}{100} = \frac{1}{2} \\ \overline{xy} &= \frac{0 \times 0 \times 40 + 0 \times 1 \times 10 + 1 \times 0 \times 10 + 1 \times 1 \times 40}{100} = \frac{2}{5}\end{aligned}$$

より共分散は

$$\overline{xy} - \bar{x} \times \bar{y} = \frac{3}{20}$$

となります。次に,

$$\begin{aligned}\overline{(x^2)} &= \frac{0^2 \times 40 + 0^2 \times 10 + 1^2 \times 10 + 1^2 \times 40}{100} = \frac{1}{2} \\ \overline{(y^2)} &= \frac{0^2 \times 40 + 1^2 \times 10 + 0^2 \times 10 + 1^2 \times 40}{100} = \frac{1}{2}\end{aligned}$$

より, x, y の分散は

$$\begin{aligned}V[x] &= \overline{(x^2)} - (\bar{x})^2 = \frac{1}{4} \\ V[y] &= \overline{(y^2)} - (\bar{y})^2 = \frac{1}{4}\end{aligned}$$

となります。よって相関係数は,

$$\frac{3/20}{\sqrt{1/4}\sqrt{1/4}} = \frac{3}{5}$$

となります。

7の正解は①となります。

[3]

属性 B のデータをすべて $1/2$ 倍しても、相関係数の値は変わりません。よってまず $r_1 = r_2$ です。

次に、 r_3 について、今度は $(1, 0)40$ 個、 $(0, 0)10$ 個、 $(1, 1)10$ 個、 $(0, 1)40$ 個の相関係数を求めることとなります。 \bar{x} , \bar{y} , $V[x]$, $V[y]$ は r_1 のときと変わりませんが、 \overline{xy} のみ $1/10$ となります（上記と同様に計算するとわかります）。すると、共分散は $1/10 - (1/2)^2 = -3/20$ となり、 r_1 と符号が逆になります。

以上より、8の正解は①となります。

問7（解答番号 16～18）

[1]

$$P(A \text{ が } B \text{ に勝つ}) \times P(A \text{ が } C \text{ に負ける}) \times P(A \text{ が優勝} \mid A \text{ が負ける}) = \frac{1}{2} \times \frac{1}{2} \times r = \frac{1}{4}r$$

です。

16の正解は⑤となります。

[2]

「[1] の場合の確率」: $\frac{1}{4}r$

「最初に A が 2 連勝する確率」: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

「最初 B に負け、そのあと A が優勝する確率」: $\frac{1}{2} \times r$

これらは互いに排反なので、足し合わせて、 $P_A = \frac{1}{4} + \frac{3}{4}r$ となります。

17の正解は③となります。

[3]

A と B は対称的なので、 $P_A = P_B$ です。また、 C に関しては、

「最初に A が B に勝ち、次に C が A に勝ち、次に B にも勝ち優勝する」: $\frac{1}{8}$

「最初に B が A に勝ち、次に C が B に勝ち、次に A にも勝ち優勝する」: $\frac{1}{8}$

「最初に A が B に勝ち、次に C が A に勝ち、次に C が B に負け、その後優勝する」: $\frac{1}{8}r$

「最初に B が A に勝ち、次に C が B に勝ち、次に C が A に負け、その後優勝する」: $\frac{1}{8}r$

これらは互いに排反なので、足し合わせて、 $P_C = \frac{1}{4} + \frac{1}{4}r$ となります。これは P_A , P_B より小さいです。

18の正解は②となります。

問 8 (解答番号 19 ~ 20)

[1]

中央値は、 $F(x) = 0.5$ のところですので、 $x^2 = 0.5$ を解いて、0.707 となります。

19 の正解は④となります。

[2]

確率密度関数 $f(x)$ は、分布関数 $F(x)$ を x で微分することで得られます。 $0 \leq x \leq 1$ では、 $f(x) = 2x$ で、それ以外では $f(x) = 0$ です。

$$E[X] = \int_0^1 x \times 2x dx = \left[\frac{2}{3} x^3 \right]_0^1 = \frac{2}{3}$$
$$E[X^2] = \int_0^1 x^2 \times 2x dx = \left[\frac{1}{2} x^4 \right]_0^1 = \frac{1}{2}$$

より、

$$V[X] = E[X^2] - (E[X])^2 = \frac{1}{18} \approx 0.056$$

となります。20 の正解は②となります。

問 12 (解答番号 29)

まず第 2 種の過誤率を求めます。第 2 種の過誤率は、「 H_1 が正しいときに、誤って H_0 を採択する確率」です。

いま、帰無仮説 H_0 : 「 $\mu = 600$ 」, 対立仮説 H_1 : 「 $\mu = 630$ 」とし、 H_1 が正しいとします。

ところが検定を行う者はそんなことはわからず、機械的に、 $\frac{\bar{X} - 600}{10} < 2.33$ なら、 H_0 を採択してしまいます。ここで、 \bar{X} は正規分布 $N(630, 50^2/25)$ に従いますので、この確率を求めると、

$$\begin{aligned} & P\left(\frac{\bar{X} - 600}{10} < 2.33\right) \\ &= P(\bar{X} < 623.3) \\ &= P\left(\frac{\bar{X} - 630}{\sqrt{50^2/25}} < \frac{623.3 - 630}{\sqrt{50^2/25}}\right) \\ &= P(Z < -0.67) \quad (Z \text{ は標準正規分布に従う確率変数}) \\ &= 0.2514 \end{aligned}$$

となります。これが第 2 種の過誤率です。検出力 = $1 -$ 第 2 種の過誤率 ですので、検出力は約 0.75 となります。

29 の正解は④となります。

【第7回：2015年11月実施】

問11 (解答番号 21 ~ 22)

[1]

1000歩中のカウントされた歩数を S と表しますと、 S は二項分布 $B(1000, 1/10)$ に従います。二項近似を行いますと、 S は正規分布 $N\left(1000 \times \frac{1}{10}, 1000 \times \frac{1}{10} \times \frac{9}{10}\right)$, つまり $N(100, 90)$ に従います。よって

$$P(S \geq 110) = P\left(\frac{S - 100}{\sqrt{90}} \geq \frac{110 - 100}{\sqrt{90}}\right) \approx P(Z > 1.054) \approx 0.15$$

となります。21 の正解は②となります。

[2]

「カウント数が67である」という事象を A , 「600歩あるく」という事象を B , 「700歩あるく」という事象を C としますと、

$$\begin{aligned} P(B | A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | C)P(C)} \\ &= \frac{\frac{1}{2} \times {}_{600}C_{67} \left(\frac{1}{10}\right)^{67} \left(\frac{9}{10}\right)^{600-67}}{\frac{1}{2} \times {}_{600}C_{67} \left(\frac{1}{10}\right)^{67} \left(\frac{9}{10}\right)^{600-67} + \frac{1}{2} \times {}_{700}C_{67} \left(\frac{1}{10}\right)^{67} \left(\frac{9}{10}\right)^{700-67}} \end{aligned}$$

となり、 ${}_{600}C_{67} \left(\frac{1}{10}\right)^{67} \left(\frac{9}{10}\right)^{600-67} = 0.033$, ${}_{700}C_{67} \left(\frac{1}{10}\right)^{67} \left(\frac{9}{10}\right)^{700-67} = 0.048$ となるのですが、これを試験時間内に解くのは大変難しいです。

そこで、正規近似を行います。

$$\begin{aligned} P_1 &= {}_{600}C_{67} \left(\frac{1}{10}\right)^{67} \left(\frac{9}{10}\right)^{600-67} \\ &\approx P(66.5 < S < 67.5) \quad (S \text{ はこのとき二項分布 } B(600, 1/10) \text{ に従っている}) \\ &= P\left(\frac{66.5 - 600 \times 1/10}{\sqrt{600 \times 1/10 \times 9/10}} < Z < \frac{67.5 - 600 \times 1/10}{\sqrt{600 \times 1/10 \times 9/10}}\right) \\ &\approx P(0.88 < Z < 1.02) = 0.1894 - 0.1539 = 0.0355 \end{aligned}$$

$$\begin{aligned} P_2 &= {}_{700}C_{67} \left(\frac{1}{10}\right)^{67} \left(\frac{9}{10}\right)^{700-67} \\ &\approx P(66.5 < S < 67.5) \quad (S \text{ はこのとき二項分布 } B(700, 1/10) \text{ に従っている}) \\ &= P\left(\frac{66.5 - 700 \times 1/10}{\sqrt{700 \times 1/10 \times 9/10}} < Z < \frac{67.5 - 700 \times 1/10}{\sqrt{700 \times 1/10 \times 9/10}}\right) \\ &\approx P(-0.44 < Z < -0.31) = 0.3783 - 0.33 = 0.0483 \end{aligned}$$

となります。よって、求める答えは、

$$\frac{\frac{1}{2} \times 0.0355}{\frac{1}{2} \times 0.0355 + \frac{1}{2} \times 0.0483} \approx 0.42$$

となります。[22]の正解は③となります。

問 13 (解答番号 [24])

I : 正しいです (以下の定義も参照)。

II : 正しいです。 $E[X_1] = \mu$ ですが一致推定量ではありません。

III : まずは、以下の一致性の定義などをご覧ください。

一致性 : 母数 θ の推定量 $\hat{\theta}$ について、標本数 n が大きくなるにつれ、任意の正の数 $\varepsilon (> 0)$ に対して、

$$P(|\hat{\theta} - \theta| > \varepsilon) \rightarrow 0$$

が成り立つとき、 $\hat{\theta}$ は母数 θ の一致推定量である、という。

なお、標本平均 \bar{X} は μ の一致推定量である。このことは、チェビシエフの不等式を用いて示すことができる。母分散を σ^2 と表すと、 $E[\bar{X}] = \mu$ 、分散 $V[\bar{X}] = \frac{\sigma^2}{n}$ であるから、

$$P\left(|\bar{X} - E[\bar{X}]| > k\sqrt{V[\bar{X}]}\right) < \frac{1}{k^2} \iff P\left(|\bar{X} - \mu| > \frac{k\sigma}{\sqrt{n}}\right) < \frac{1}{k^2}$$

となるが、ここで $\frac{k\sigma}{\sqrt{n}} = \varepsilon$ とおくと、

$$P(|\bar{X} - \mu| > \varepsilon) < \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \quad (n \rightarrow \infty)$$

となり、確かに $n \rightarrow \infty$ において右辺は 0 に収束し、標本平均 \bar{X} は μ の一致推定量になる。

さて、例えば $\bar{X}' = \frac{3X_1 - X_2 + X_3 + X_4 + \dots + X_n}{n}$ という推定量を考えますと、この期待値は μ になり、分散は $\frac{n+8}{n^2}\sigma^2$ となります。上のやり方と同じように k を置くことで、一致性をいうことができます。

したがって、III は誤りです。

[24]の正解は②となります。

問 17 (解答番号 [31])

第 1 種の過誤率は、 H_0 が正しいときに H_0 を棄却する条件付確率です。

「 H_0 が正しい」つまり「 $p = 0.62$ 」

のときに、

「 H_0 を棄却する」つまり「3 回とも針が上向きか、3 回とも下向き（これらは互いに排反）」

となる確率なので、②となります。

31 の正解は②となります。

【第6回：2015年6月実施】

問8 (解答番号 16 ~ 18)

[1]

$$\frac{5 \times 7 \times 3}{{}_{15}C_3} = \frac{3}{13}$$

となり、16の正解は⑤となります。

[2]

100円玉が2枚と、あと1枚(何でもよい)あればよいです。100円玉3枚をとる場合、100円玉2枚と10円玉1枚をとる場合、100円玉2枚と1円玉1枚をとる場合をそれぞれ考えて、

$$\frac{{}_5C_3 + {}_5C_2 \times 7 + {}_5C_2 \times 3}{{}_{15}C_3} = \frac{22}{91}$$

となります。17の正解は④となります。

[3]

取り出した3枚の合計が150円以上で、かつ1円玉が含まれている確率は、 $\frac{{}_5C_2 \times 3}{{}_{15}C_3}$ となります。この確率を[2]の確率で割ると、 $\frac{3}{11}$ となり、これが答えです。18の正解は②となります。

問11 (解答番号 23)

二項分布を用いて解くと計算が大変なことになりますので、 $n = 36$ と試行回数が小さいのですが、正規近似をします。正解した数を S と表しますと、 S は二項分布 $B(36, 1/2)$ に従い、ここで正規近似を行いますと、 S は期待値 $36 \times \frac{1}{2} = 18$ 、分散 $36 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = 9$ の正規分布に従います。よって、

$$P(S \geq 24) = P\left(\frac{S - 18}{3} \geq \frac{24 - 18}{3}\right) = P(Z \geq 2) = 0.0228$$

となります (Z は標準正規分布に従う確率変数)。

ここで、試行回数が小さいので、連続補正を行いますと、

$$P(S \geq 24 - 0.5) = P(Z \geq 1.83) = 0.0336$$

となります。なお、実際に二項分布のまま計算もしてみましたが、約3%になりました(試験時間内で計算するのは難しい)。23の正解は⑤となります。

問 13 (解答番号 25 ~ 26)

[1]

母分散未知の場合の母平均の検定です。25の正解は③となります。

[2]

第二種の過誤率とは、 H_1 が正しいときに H_0 を採択してしまう確率です。いま、 H_1 が何であれ、 H_0 を採択してしまう確率は、一様乱数の値が 0.01 より大きい確率なので、0.99 になります。

なお、第一種の過誤率とは、 H_0 が正しいときに H_0 を棄却する確率であり、有意水準に一致しますので、いまは 0.01 になり、不合理な値ではありません。

26の正解は③となります。

【第5回：2014年11月実施】

問6 (解答番号 8 ~ 11)

[1]

以下のデータの平均を求めることになります。

$$\underbrace{0, 0, 0, \dots, 0}_{55 \text{ 個}}, \underbrace{1, 1, \dots, 1}_{144 \text{ 個}}, \underbrace{2, 2, \dots, 2}_{140 \text{ 個}}, \dots, \underbrace{6, 6, 6, 6, 6}_{6 \text{ 個}}$$

よって,

$$(0 \times 55 + 1 \times 144 + 2 \times 140 + \dots + 6 \times 6) \div 500 = 2.00$$

となります。8の正解は③です。

[2]

母数 λ のポアソン分布の確率関数は,

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

です。9の正解は⑤です。

[3]

母数 λ のポアソン分布の期待値 $E[X]$, 分散 $V[X]$ はともに λ です。ここで $V[X] = E[X^2] - (E[X])^2$ ですので,

$$E[X^2] = V[X] + (E[X])^2 = \lambda + \lambda^2$$

になります。10の正解は④です。

[4]

検定統計量としては,

$$\chi^2 = \frac{(55 - 67.7)^2}{67.7} + \frac{(144 - 135.3)^2}{135.3} + \dots + \frac{(6 - 8.3)^2}{8.3} \approx 4.498$$

を用い, 自由度 5 の χ^2 分布の上側%点と比較します。

自由度に関して, セルが 7 つあるので, $7 - 1 = 6$ とし, さらに, 母数を 1 つ最尤推定値で推定しているため, 1 減らして, 5 とします (推定した母数の数だけ自由度を減じます)。

以上より, 11の正解は⑤です。

問 12 (解答番号 32 ~ 35)

[1]

やや直感に頼るところもあります。32 の正解は④になります。

[2-1]

決定係数が該当します。33 の正解は①になります。

[2-2]

母回帰方程式

$$Y = \alpha + \beta X + \varepsilon$$

において、誤差項 ε が標本毎に独立に正規分布 $N(0, \sigma^2)$ (σ^2 は定数) に従うとしますと、回帰係数の推定量 (標本から得られる回帰係数) $\hat{\alpha}$, $\hat{\beta}$ は正規分布に従います。

本問の場合、詳細は非常に長くなりますので割愛しますが、男子、女子それぞれの回帰係数の推定量 $\hat{\beta}^{(1)}$, $\hat{\beta}^{(2)}$ は正規分布に従います。すると、 $\hat{\beta}^{(1)} - \hat{\beta}^{(2)}$ も正規分布に従います。このことを利用して、それぞれの母回帰係数について「 $\beta^{(1)} = \beta^{(2)}$ 」を検定することができます (実際は t 分布を用います)。

34 の正解は②になります。

[3]

本問における重回帰モデルは、体重 Y 、身長 X 、性別 Z (男性ならば $Z = 1$ 、女性ならば $Z = 0$) と表しますと、

$$Y = \alpha + \beta^{(1)}X + \beta^{(2)}Z + \beta^{(3)}XZ + \varepsilon \cdots (a)$$

となっています。

I は、誤りです。本問のように、ダミー変数として用いて、意味のある分析ができます。

II について、(a) の式で $Z = 0$ としますと、回帰直線 $Y = \alpha + \beta^{(1)}X$ が現れます。II は誤りです。偶然ではありません。

III について、男性のデータを用いる、つまり (a) で $Z = 1$ としますと、回帰直線 $Y = \alpha + \beta^{(2)} + (\beta^{(1)} + \beta^{(3)})X$ が現れます。女性のデータを用いる、つまり (a) で $Z = 0$ としますと、回帰直線 $Y = \alpha + \beta^{(1)}X$ が現れます。

$\beta^{(3)}$ は、男性と女性の回帰係数の差に対応しているといえます。III は正しいです。

以上より、35 の正解は③になります。

【第4回：2014年6月実施】

問13（解答番号 20～22）

[1]

3の目が出る回数 r は二項分布 $B(7, 1/6)$ に従い、

$${}^7C_4 \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^3$$

が求める確率になります。 ${}^7C_4 = 35$ なので、20の正解は④となります。

[2]

帰無仮説が正しいとしたときに、「3の目が4回出る」という事象が起こる確率が非常に小さければ、滅多に起こらないことが起こったということで、帰無仮説を疑うこととなります。このとき、5回以上の確率も加えた確率を P 値とし、有意水準である5%と比較します。

21の正解は①となります。

[3]

1の目が4回出る確率も[1]と同じ式で計算できて0.015629、2の目が4回出る確率も同様に0.015629、…、です。これらの事象は互いに排反なので、求める確率は、 $0.015629 \times 6 \approx 0.094$ となります。

22の正解は③となります。

問18（解答番号 32～33）

[1]

標本回帰方程式によると、気温が1度上がると傾きの大きさだけ（4.8617）降水量が増えます。

32の正解は④となります。

[2]

決定係数はB君の分析の方が大きいので、Iは誤りです。

切片や平均気温の標準誤差をみると、B君の分析の方が小さくなっています。これらは、標本回帰方程式における切片と傾きの標準偏差です。IIは正しいです。

自由度修正済み決定係数は、説明変数の数の検討に用いられます。単回帰分析同士のこの決定係数の比較は意味がありません。IIIは誤りです。

以上より、33の正解は②となります。

【第3回：2013年11月実施】

問4 (解答番号 5)

相関係数は、各変数を標準化しても値は変わりませんので、 A と C の値は同じになります。なお、 D は、 x, y を標準化したものをそれぞれ x', y' とおくと、

$$x'_i = \frac{x_i - \bar{x}}{s_x}, \quad y'_i = \frac{y_i - \bar{y}}{s_y}$$

であり、 $\bar{x}' = \bar{y}' = 0$ です。よって、

$$\begin{aligned} \text{Cov}(x', y') &= \frac{1}{n-1} \sum (x'_i - \bar{x}')(y'_i - \bar{y}') \\ &= \frac{1}{n-1} \sum x'_i \cdot y'_i \\ &= \frac{1}{n-1} \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y} \end{aligned}$$

となりますが、最後の式に問題文にある s_x と s_y を代入すると、 r_{xy} と等しくなることがわかります。

以上より、5の正解は⑤となります。

なお、本問では分散、共分散ともに不偏推定量を用いています。不偏でない分散・共分散 ($n-1$ で割らずに n で割るもの) か、それとも本問のような分散・共分散か、テキストや試験によってどちらを使うかマチマチですので、問題文をよく読んで間違えないようにして下さい (統計検定試験は不偏推定量で統一されています)。

問16 (解答番号 25 ~ 26)

[1]

標本比率 \hat{p} 、標本数 n と表すと、母比率 p の 95% 信頼区間は、

$$\hat{p} \pm 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

で表されます。 $\hat{p} = 0.25$ 、 $n = 600$ なので、25の正解は⑤となります。

[2] 信頼区間の幅は、

$$2 \times 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ですので、

$$2 \times 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.04$$

すなわち

$$n \geq \left(\frac{2 \times 1.96}{0.04} \right)^2 \times \hat{p}(1 - \hat{p})$$

となります。 $\hat{p}(1 - \hat{p})$ は $0 \leq \hat{p} \leq 0.3$ において単調増加なので、 $\hat{p} = 0.3$ のときに最大になりますので、

$$n \geq \left(\frac{2 \times 1.96}{0.04} \right)^2 \times (0.3 \times 0.7) = 2016$$

でなければなりません。26の正解は④となります。

【第2回：2012年11月実施】

問13 (解答番号 20)

n 個の世帯について、各世帯の人数を X_1, X_2, \dots, X_n と表しますと、「世帯毎の人数」の推定量は $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ となります。すると、人口の推定量は $100000\bar{X}$ となります。その変動係数は、 $E[\bar{X}] = E[X_i]$, $V[\bar{X}] = \frac{V[X_i]}{n}$ などより、

$$\frac{\sqrt{V[100000\bar{X}]}}{E[100000\bar{X}]} = \frac{1}{\sqrt{n}} \frac{\sqrt{V[X_i]}}{E[X_i]}$$

となります。

ところで、母集団の各世帯における人数の母平均を μ , 母分散を σ^2 と表しますと、

$$E[X_i] = \mu, \quad V[X_i] = \sigma^2$$

ですので、上記の変動係数は

$$\frac{1}{\sqrt{n}} \frac{\sigma}{\mu}$$

となりますが、ここで母集団の変動係数（下線部）が 1.0 以下なので、

$$\frac{1}{\sqrt{n}} \leq 0.05$$

より $n \geq 400$ となります。20 の正解は②となります。

問 14 (解答番号 21 ~ 22)

[1]

入場者数を N 人としますと、 i 番目の人の通し番号 X_i は区間 $(1, N)$ の離散型一様分布に従います。その期待値は $E[X_i] = \frac{N+1}{2}$ です。よって、通し番号の合計 $S = X_1 + X_2 + \dots + X_{300}$ の期待値は $E[S] = 300 \times \frac{N+1}{2}$ です。いま、 N の推定量を \hat{N} と表し、

$$\hat{N} = 2 \times \frac{S}{300} - 1$$

としますと、 $E[\hat{N}] = N$ となり、 N の不偏推定量になります。よって、 $2 \times \frac{4510384}{300} - 1 \approx 30000$ が答えとなります。21 の正解は③となります。

[2]

N の標準誤差とは、 \hat{N} の標準偏差です。

$$V[\hat{N}] = \frac{2^2}{300^2} \times 300V[X_i]$$

ですが、ここで $V[X_i]$ は、

$$\begin{aligned} E[X_i^2] &= \frac{1}{N}(1^2 + 2^2 + \dots + N^2) \\ &= \frac{1}{N} \times \frac{1}{6}N(N+1)(2N+1) \\ &= \frac{1}{6}(N+1)(2N+1) \end{aligned}$$

より、

$$\begin{aligned} V[X_i] &= E[X_i^2] - (E[X_i])^2 \\ &= \frac{1}{6}(N+1)(2N+1) - \frac{1}{4}(N+1)^2 \\ &= \frac{1}{12}(N-1)(N+1) \end{aligned}$$

となります。 $N \approx 30000$ を代入して、 $\sqrt{V[\hat{N}]} \approx 1000$ を得ます。22 の正解は②となります。

なお、 N が非常に大きいことから、離散型の一様分布ではなく連続型の一様分布とみなし、

$$V[X_i] = \frac{(N-1)^2}{12}$$

として、 $N \approx 30000$ を代入しても同じ答えが得られます。

【第1回：2011年11月実施】

問5 (解答番号 9)

30匹の小型犬の体重を x_1, x_2, \dots, x_{30} , それらの相加平均を \bar{x} , 標準偏差を s と表しますと,

$$\bar{x} = \frac{1}{30} \sum_{i=1}^{30} x_i = 4.5$$
$$s^2 = \frac{1}{30} \sum_{i=1}^{30} (x_i - \bar{x})^2 = 1.0^2$$

であり, 変動係数 $a = s/\bar{x} = 1.0/4.5$ となります。ここで, 各犬の体重が 0.5kg ずつ増加したとして, それら新たな体重を $x'_1, x'_2, \dots, x'_{30}$ と表しますと, $x'_i = x_i + 0.5$ であり, 新たな相加平均を \bar{x}' , 標準偏差を s' と表しますと, $\bar{x}' = \bar{x} + 0.5$, $s = s'$ が成り立ちます。よって,

$$(\text{新たな変動係数}) = s'/\bar{x}' = s/(\bar{x} + 0.5) = 1.0/5.0 = \frac{5.0}{4.5}a = \frac{10}{9}a$$

となります。よって, 9 の正解は3となります。

問14 (解答番号 18)

平均 μ , 分散 σ^2 の正規分布の確率密度関数 $f(x)$ は,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

です。a. $x = \mu$ において $f(x)$ は最大値 $\frac{1}{\sqrt{2\pi}\sigma}$ をとります。b. σ が大きいほど, その最大値は大きいです。よって, a. と b. は正しいです。c. X を $N(0, \sigma_1^2)$ に従う確率変数, Y を $N(0, \sigma_2^2)$ に従う確率変数, Z を標準正規分布に従う確率変数としますと, $Pr(X \geq x) = Pr(Z \geq x/\sigma_1)$, $Pr(Y \geq x) = Pr(Z \geq x/\sigma_2)$ と書き直せます。 $x \geq 0$, $\sigma_1 < \sigma_2$ から $x/\sigma_1 \geq x/\sigma_2$ なので, $Pr(Z \geq x/\sigma_1) \leq Pr(Z \geq x/\sigma_2)$, すなわち $Pr(X \geq x) \leq Pr(Y \geq x)$ です。よって c. も正しいです。

以上より, 18 の正解は5となります。

問17 (解答番号 21)

母支持率を p , 標本支持率を \hat{p} , 標本数を n と表しますと, 通常 n は非常に大きいので, 中心極限定理を用いて, \hat{p} は正規分布 $N\left(p, \frac{p(1-p)}{n}\right)$ に従うと近似できます。よって, 例えば p の 95% 信

信頼区間は,

$$Pr \left(p - Z_{0.025} \sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + Z_{0.025} \sqrt{\frac{p(1-p)}{n}} \right) = 0.95$$

の () の中を, 分散に現れる p を \hat{p} で近似し, そののち p について整理した,

$$\hat{p} - Z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + Z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

となります。この信頼区間の幅は, $2Z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ です。 n が 4 倍になれば, この幅は半分になることがわかります。信頼係数が変わっても一緒です。

以上より, 21 の正解は 4 となります。

(以上です)