

問題 1 厚生労働省は、2020 年 6 月 1 日～6 月 7 日にかけて東京都・大阪府・宮城県において無作為抽出された一般住民を対象に、新型コロナウイルス抗体検査を実施しました。大阪府では観測数 2970 名に対して抗体保有者が 5 名となりました。大阪府における抗体保有率 p の 95 %信頼区間を求めて下さい。

ただし母集団（大阪府の一般住民全体）からとった 2970 個の標本中の抗体保有者数を X とし、 X の分散をサンプル値を使って

$$2970p(1-p) \xrightarrow{\text{代用}} 2970 \cdot \frac{5}{2970} \cdot \frac{2965}{2970} \approx 2.234^2$$

のように 2.234^2 で代用したうえで X を正規分布 $N(2970p, 2.234^2)$ によって近似し、半整数補正も行して下さい。

配点：10 点 シラバス到達目標：カ、キ、ク

【解答例】 X は 2 項分布 $B(2970, p)$ に従いますが、指示の通りこれを正規分布 $N(2970p, 2.234^2)$ によって近似して

$$P[|X - 2970p| \leq d] = 0.95$$

となる $d > 0$ を求めます。

$$\begin{aligned} 0.95 &= P[|X - 2970p| \leq d] \\ &\approx P[|N(2970p, 2.234^2) - 2970p| \leq d + 0.5] \\ &= P[|N(0, 2.234^2)| \leq d + 0.5] \\ &= P\left[|N(0, 1)| \leq \frac{d + 0.5}{2.234}\right] \\ 0.475 &= P\left[0 \leq N(0, 1) \leq \frac{d + 0.5}{2.234}\right] \end{aligned}$$

から、正規分布表によれば

$$\frac{d + 0.5}{2.234} \approx 1.96 \quad \text{すなわち} \quad d \approx 3.879$$

が分かりました。

従って今回のサンプル値 5 について、信頼度 95 %で

$$|5 - 2970p| \leq 3.879$$

$$1.121 \leq 2970p \leq 8.879$$

$$0.000377 \leq p \leq 0.00299$$

となりますから求める信頼区間は $[0.0377\%, 0.299\%]$ です。

【半整数補正をしない場合】正規分布表によれば

$$\frac{d}{2.234} \approx 1.96 \quad \text{すなわち} \quad d \approx 4.379$$

です。

従って今回のサンプル値 5 について、信頼度 95 %で

$$|5 - 2970p| \leq 4.379$$

$$0.621 \leq 2970p \leq 9.379$$

$$0.000209 \leq p \leq 0.00316$$

となりますから求める信頼区間は $[0.0209\%, 0.316\%]$ です。

問題 2 健常者の血清カルシウムの平均値は 9.8mg/dl です。無作為に選んだ副甲状腺機能低下症の患者 16 人の血清カルシウム値を測定したところ、平均値は 7.4mg/dl でした。副甲状腺機能低下症患者的血清カルシウム平均値 m は健常者に比べて低いと言えるかどうか有意水準 0.01 で仮説検定して下さい。ただし、副甲状腺機能低下症患者的血清カルシウム値全体は、標準偏差 0.5mg/dl の正規分布に従うと仮定します。

配点：10 点 シラバス到達目標：カ、キ、ケ

【解答例】

帰無仮説 H_0 : 副甲状腺機能低下症患者的血清カルシウム平均値は 9.8mg/dl である。

帰無仮説が正しいと仮定すると、母集団、すなわち副甲状腺機能低下症患者的血清カルシウム値全体は正規分布 $N(9.8, 0.5^2)$ に従うので、ここから取った大きさ 16 の標本平均 $\bar{X}$ は正規分布 $N\left(9.8, \frac{0.5^2}{16}\right)$ に従います。

そこで有意水準 0.01 での左片側棄却域を求めるために

$$0.01 = P[\bar{X} \leq 9.8 - d]$$

となるような $d > 0$ を求めます。

$$\begin{aligned} 0.01 &= P[\bar{X} \leq 9.8 - d] \\ &= P\left[N\left(9.8, \frac{0.5^2}{16}\right) \leq 9.8 - d\right] \end{aligned}$$

$$= P[N(0, 1) \leq -8d]$$

$$0.49 = P[0 \leq N(0, 1) \leq 8d]$$

なので標準正規分布表によれば

$$8d \approx 2.33, \quad \text{すなわち} \quad d \approx 0.291$$

が得られ、求める左片側棄却域は $(-\infty, 9.509]$ となります。

今回のサンプル平均 7.4 はこの棄却域に入っていますので、帰無仮説は棄却され、副甲状腺機能低下症患者的血清カルシウム平均値は健常者に比べて低いと判断されます。

【表の読み方のヴァリエーション】もしくは

$$8d \approx 2.325, \quad \text{すなわち} \quad d \approx 0.291$$

と読んだ場合は、左片側棄却域は $(-\infty, 9.509]$ となり、同様に帰無仮説は棄却されます。

【両側検定でやってしまった場合】有意水準 0.01 での両側棄却域を求めるために

です。

$$0.01 = P[|\bar{X} - 9.8| \geq d]$$

となるような $d > 0$ を求めます。

$$\begin{aligned} 0.01 &= P[|\bar{X} - 9.8| \geq d] \\ &= P\left[\left|N\left(9.8, \frac{0.5^2}{16}\right) - 9.8\right| \geq d\right] \\ &= P[|N(0, 1)| \geq 8d] \\ 0.495 &= P[0 \leq N(0, 1) \leq 8d] \end{aligned}$$

なので標準正規分布表によれば

$$8d \approx 2.575, \quad \text{すなわち} \quad d \approx 0.322$$

が得られ、求める両側棄却域は $(-\infty, 9.478] \cup [10.122, \infty)$ となります。

(以上いずれの場合も) 今回のサンプル平均 7.4 はこの棄却域に入っていますので、帰無仮説は棄却され、副甲状腺機能低下症患者の血清カルシウム平均値は健常者とは異なると判断されます。 □

もしくは表の読み方のヴァリエーションとしては

$$8d \approx 2.57, \quad \text{すなわち} \quad d \approx 0.321$$

あるいは

$$8d \approx 2.58, \quad \text{すなわち} \quad d \approx 0.323$$

とする場合があります、棄却域はそれぞれ

$$(-\infty, 9.479] \cup [10.121, \infty), \quad (-\infty, 9.477] \cup [10.123, \infty)$$

問題 3 次の 2 次元データ (X, Y) において：

X	0.75	0.55	0.72	0.62	0.66
Y	8.7	6.8	7.9	7.0	7.1

X, Y それぞれの平均値と X, Y の共分散を求めて下さい（近似はしないこと）。

配点：30 点 シラバス到達目標：ア、イ、オ

【解答例】

$$E[X] = \frac{0.75 + \cdots + 0.66}{5} = \frac{3.3}{5} = 0.66$$

$$E[Y] = \frac{8.7 + \cdots + 7.1}{5} = \frac{37.5}{5} = 7.5$$

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= \frac{0.09 \cdot 1.2 + 0.11 \cdot 0.7 + 0.06 \cdot 0.4 + 0.04 \cdot 0.5 + 0}{5} \\ &= \frac{0.108 + 0.077 + 0.024 + 0.02}{5} \\ &= \frac{0.229}{5} \\ &= 0.0458 \end{aligned}$$

□

問題 4 ある地域に住む 25 歳の女性の中から 500 人を抽出して身長を測定したところ、平均値が 157.6、不偏分散が 4.98^2 でした。この地域における 25 歳の女性の身長の平均値の信頼区間を信頼度 95 % で求めて下さい。

配点：20 点 シラバス到達目標：カ、キ、ク

【解答例】 この地域の 25 歳女性の身長データ全体を母集団 X とし、その平均値を m とします。この母集団から取った大きさ 500 の標本平均を $\bar{X}$ とします。

まず母分散をサンプルの不偏分散 4.98^2 で代用し、さらに中心極限定理によれば $\bar{X}$ は正規分布 $N\left(m, \frac{4.98^2}{500}\right)$ で近似されます。

そこで

$$P\left[|\bar{X} - m| \leq d\right] = 0.95$$

となる d を求めると、

$$\begin{aligned} 0.95 &\approx P\left[\left|N\left(m, \frac{4.98^2}{500}\right) - m\right| \leq d\right] \\ &= P\left[\left|N\left(0, \frac{4.98^2}{500}\right)\right| \leq d\right] \\ &= P\left[\left|N(0, 1)\right| \leq \frac{d}{\sqrt{\frac{4.98^2}{500}}}\right] \end{aligned}$$

$$0.475 = P\left[0 \leq N(0,1) \leq \frac{d}{\sqrt{\frac{4.98^2}{500}}}\right]$$

であり、標準正規分布表から

$$\begin{aligned} \frac{d}{\sqrt{\frac{4.98^2}{500}}} &\approx 1.96 \\ d &\approx 1.96 \cdot \frac{4.98}{22.36} \approx 0.4365 \end{aligned}$$

が得られます。

従って今回のサンプル平均 157.6 について、信頼度 95 %で

$$\begin{aligned} |157.6 - m| &\leq 0.4365 \\ 157.1635 &\leq m \leq 158.0365 \end{aligned}$$

が成り立ちますから求める信頼区間は [157.16, 158.04] です。

□

問題 5 プロサッカーリーグに所属する B チーム、L チームの 1 試合あたりの得点数 G_B, G_L はそれぞれパラメータ 1、2 のポアソン分布に従い、互いに独立であると仮定します。

今回 B チームと L チームが対戦します。B チームは 1 試合当たり平均 1 得点ですから、平均 2 得点の L チームに比べて劣勢ですが、『番狂せ』を起こす確率を以下のように計算してみましょう。

- (1) 『1 対 0 で B チームが勝つ』確率、 $P[G_B = 1, G_L = 0]$ を求めて下さい。
- (2) $P[G_B = 2, G_L = 0] + P[G_B = 2, G_L = 1]$ を求めて下さい。
- (3) B チームが 4 点以内の得点で L チームに勝つ確率を求めて下さい。

配点：(1)4 点、(2)4 点、(3)2 点	シラバス到達目標：エ、オ
-------------------------	--------------

【解答例】 (1)

$$P[G_B = 1, G_L = 0] = P[G_B = 1]P[G_L = 0] = \frac{1^1}{1!}e^{-1}\frac{2^0}{0!}e^{-2} = e^{-3} \approx 0.0498$$

(2)

$$\begin{aligned} &P[G_B = 2, G_L = 0] + P[G_B = 2, G_L = 1] \\ &= P[G_B = 2]P[G_L = 0] + P[G_B = 2]P[G_L = 1] \\ &= \frac{1^2}{2!}e^{-1}\left(\frac{2^0}{0!} + \frac{2^1}{1!}\right)e^{-2} \\ &= \frac{3}{2}e^{-3} \\ &\approx 0.0747 \end{aligned}$$

(3)

$$\begin{aligned}
 P[G_B = 3, G_L \leq 2] &= P[G_B = 3] (P[G_L = 0] + P[G_L = 1] + P[G_L = 2]) \\
 &= \frac{1^3}{3!} e^{-1} \left(1 + 2 + \frac{2^2}{2!} \right) e^{-2} \\
 &= \frac{5}{6} e^{-3} \\
 &\approx 0.0415
 \end{aligned}$$

(2)

$$\begin{aligned}
 &P[G_B = 2, G_L = 0] + P[G_B = 2, G_L = 1] \\
 &= P[G_B = 2] P[G_L = 0] + P[G_B = 2] P[G_L = 1] \\
 &= \frac{1^2}{2!} e^{-1} \left(\frac{2^0}{0!} + \frac{2^1}{1!} \right) e^{-2} \\
 &= \frac{3}{2} e^{-1} \times e^{-2} \\
 &\approx 0.07452
 \end{aligned}$$

$$\begin{aligned}
 P[G_B = 4, G_L \leq 3] &= P[G_B = 4] \sum_{j=0}^3 P[G_L = j] \\
 &= \frac{1^4}{4!} e^{-1} \left(1 + 2 + 2 + \frac{2^3}{3!} \right) e^{-2} \\
 &= \frac{19}{72} e^{-3} \\
 &\approx 0.01314
 \end{aligned}$$

(3)

$$\begin{aligned}
 P[G_B = 3, G_L \leq 2] &= P[G_B = 3] (P[G_L = 0] + P[G_L = 1] + P[G_L = 2]) \\
 &= \frac{1^3}{3!} e^{-1} \left(1 + 2 + \frac{2^2}{2!} \right) e^{-2} \\
 &= \frac{5}{6} e^{-1} \times e^{-2} \\
 &\approx 0.0414
 \end{aligned}$$

以上により、求める確率は

$$0.0498 + 0.0747 + 0.0415 + 0.01314 \approx 0.17914$$

です。

$$\begin{aligned}
 P[G_B = 4, G_L \leq 3] &= P[G_B = 4] \sum_{j=0}^3 P[G_L = j] \\
 &= \frac{1^4}{4!} e^{-1} \left(1 + 2 + 2 + \frac{2^3}{3!} \right) e^{-2} \\
 &= \frac{19}{72} e^{-1} \times e^{-2} \\
 &\approx 0.01311
 \end{aligned}$$

【 e^{-3} のところを $e^{-1} \times e^{-2} \approx 0.368 \times 0.135 \approx 0.04968$ とした場合】

(1)

$$P[G_B = 1, G_L = 0] = P[G_B = 1] P[G_L = 0] = \frac{1^1}{1!} e^{-1} \frac{2^0}{0!} e^{-2} \approx 0.368 \times 0.135 \approx 0.04968$$

以上により、求める確率は

$$0.04968 + 0.07452 + 0.0414 + 0.01311 \approx 0.17871$$

でも良い。

□

問題 6 密度関数が次の $f(x)$:

$$f(x) = \begin{cases} \frac{1}{4} & -3 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

で与えられる確率変数 U の分散を求めて下さい。

配点：10 点

シラバス到達目標：ア、ウ

【解答例】

$$\int_{-\infty}^{\infty} x f(x) dx = \int_{-3}^1 \frac{1}{4} x dx = \frac{1}{8} [x^2]_{-3}^1 = -1$$

から平均値は -1 ですから、

$$\begin{aligned} \int_{-\infty}^{\infty} (x+1)^2 f(x) dx &= \int_{-3}^1 \frac{1}{4} (x+1)^2 dx \\ &= \left[\frac{1}{12} (x+1)^3 \right]_{-3}^1 \\ &= \frac{1}{12} (8+8) \\ &= \frac{4}{3} \end{aligned}$$

となり、分散は存在して $\frac{4}{3}$ です。

□

問題 7 2次元の確率変数 (X, Y) は X, Y が独立であって、 X は正規分布 $N(50, 16)$ に、 Y は正規分布 $N(15, 1)$ に従っています。このとき

$$P[X - 3Y \leq 2]$$

を求めて下さい。

配点：10 点

シラバス到達目標：エ、オ

【解答例】 $-3Y$ は正規分布 $N(-45, 9)$ に従います。

X と $-3Y$ は独立ですから、その和 $X - 3Y$ は正規分布 $N(5, 25)$ に従い、

$$\begin{aligned} P[X - 3Y \leq 2] &= P[N(5, 25) \leq 2] \\ &= P[N(0, 25) \leq -3] \\ &= P\left[N(0, 1) \leq -\frac{3}{5}\right] \\ &= 0.5 - P\left[0 \leq N(0, 1) \leq \frac{3}{5}\right] \\ &\approx 0.5 - 0.2257 \\ &= 0.2743 \end{aligned}$$

となります。

□