2021年度 大学共通テスト(1次日程) 【数学Ⅱ·数学B】解説

第3問

(1)確率0.5の事象を100回反復させるのと同じことなので、二項分布 N(100, 0.5)に従います。二項分布の平均E(X)及び分散 V(X)は、公式に従って、

$$E(X) = 100 \cdot 0.5 = 50$$
$$V(X) = 100 \cdot 0.5 \cdot 0.5 = 25$$

となります。標準偏差は、 $\sqrt{V(X)} = \sqrt{25} = 5$ です。

(2)正規分布に近似できるとされているので、以下では正規分布表を見ながら考えていきましょう。

まず、母比率を0.5とした場合の確率変数を標準化した確率変数を Z_5 とすると、(1)で平均m=50、標準偏差 $\sigma=5$ と求めているので、

$$Z_5 = \frac{X - 50}{5}$$

従って、36人以下になる確率psは、

$$p_5 = P\left(Z \le \frac{36 - 50}{5}\right) = P(Z \le -2.8)$$

正規分布表より、 $P(Z \le -2.8) = 0.5 - 0.4974 = 0.0026$ となり、小数第3位までの概数で四捨五入すると 0.003 になります。

続く、母比率を0.4とした場合の36人以下になる確率p4は、

$$E(X) = 100 \cdot 0.4 = 40$$
$$\sqrt{V(X)} = \sqrt{100 \cdot 0.4 \cdot 0.6} = 2\sqrt{6}$$

に変わるため、

$$p_4 = P\left(Z \le \frac{36 - 40}{2\sqrt{6}}\right) = P\left(Z \le -\frac{\sqrt{6}}{3}\right)$$

√6の近似値を知っているのであれば、正規分布表を用いてp₄を計算してしまう方法もありますが、p₅との大小関係を聞かれているだけなので、その計算をする必要はありません。

 $p_4 = P\left(Z \le -\frac{\sqrt{6}}{3}\right)$ であるということは、 p_4 というのは正規分布グラフの横軸の数字

が $-\frac{\sqrt{6}}{3}$ より左にある場所の面積を表しているのであり、 $p_5=P(Z\leq -2.8)$ なので、 p_5 というのは正規分布表の横軸の数字が-2. 8より左にある場所の面積を表しています。横軸の数字が大きいほど面積が大きくなることは明らかなので、 $-\frac{\sqrt{6}}{3}$ と-2. 8の大小がわかればよいわけです。

-2.8 < -2であり、 $-\frac{\sqrt{6}}{3} = -\sqrt{\frac{2}{3}} > -1$ なので、 $-2.8 < -\frac{\sqrt{6}}{3}$ となることは明らかです。従って、 $p_4 > p_5$ であることがわかります。

(3)校長先生の調査における標本平均を $\overline{X_C}$ 、母標準偏差を σ 、標本数をnとすれば、 母平均mに対する信頼度95%の信頼区間は、

$$\overline{X_C} - 1.96 \frac{\sigma}{\sqrt{n}} \le m \le \overline{X_C} + 1.96 \frac{\sigma}{\sqrt{n}}$$

で表されます。

これは覚えておくことに越したことはありませんが、1.96という数字を忘れた場合は、与えられている正規分布表から導くこともできます。信頼度95%というのは、正規分布グラフの中央部95%を占めることを意味していますので、中央から片側では95÷2=47.5%を占める必要があります。割合で考えれば 0.4750 になりますが、正規分布表からこれは $Z_0=1$.96の場合であることがわかります。そのため、 $-1.96 \le Z \le 1.96$ が信頼区間になりますが、標本平均の平均は母平均mに等しく、

標本平均の標準偏差は $\frac{\sigma}{\sqrt{n}}$ になるため、 $Z=\frac{\overline{X_C}-m}{\sigma/\sqrt{n}}$ で標準化されるので、

$$-1.96 \le \frac{\overline{X_C} - m}{\sigma / \sqrt{n}} \le 1.96$$

これをmについて解いて、上の信頼区間の公式が導かれるわけです。 信頼区間の公式と、与えられている $C_1 \le m \le C_2$ を比較すれば、

$$C_1 = \overline{X_C} - 1.96 \frac{\sigma}{\sqrt{n}}$$
, $C_2 = \overline{X_C} + 1.96 \frac{\sigma}{\sqrt{n}}$

であることがわかります。ここから、

$$C_1 + C_2 = \left(\overline{X_C} - 1.96 \frac{\sigma}{\sqrt{n}}\right) + \left(\overline{X_C} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 2\overline{X_C}$$

$$C_2 - C_1 = \left(\overline{X_C} + 1.96 \frac{\sigma}{\sqrt{n}}\right) - \left(\overline{X_C} - 1.96 \frac{\sigma}{\sqrt{n}}\right) = 2 \cdot 1.96 \frac{\sigma}{\sqrt{n}}$$

問題文から、 $\overline{X_C}=204$ 、 $\sigma=150$ 、n=100であることがわかるので、

$$C_1 + C_2 = 2 \cdot 204 = 408$$

$$C_2 - C_1 = 2 \cdot 1.96 \cdot \frac{150}{\sqrt{100}} = 58.8$$

が導かれます。

C₁はmの95%信頼区間の下限、C₂はmの95%信頼区間の上限を表しているわけですが、mが信頼区間の外側に来る可能性が完全に排除できるわけではないので、m<C₁になったりC₂<mになったりする可能性は残されています。

- (4)図書委員長の調査と校長先生の調査は独立して行われている以上、図書委員 長の調査における全く読書しなかった生徒数n人と、校長先生の調査における全く読 書しなかった生徒数36人の間の関係はわかりません。
- $(5)C_1 \ge C_2$ はそれぞれ、校長先生の調査に基づ<mの95%信頼区間の下限<上限を表しており、

$$C_1 = \overline{X_C} - 1.96 \frac{\sigma}{\sqrt{n}}$$
 , $C_2 = \overline{X_C} + 1.96 \frac{\sigma}{\sqrt{n}}$

になることは、先程示した通りです。

一方、 $D_1 \& D_2$ はそれぞれ、図書委員長の調査に基づ ζ mの95%信頼区間の下限 & E上限を表しており、図書委員長の調査における標本平均を $\overline{X_D}$ & Eとすると、母標準偏差 や標本数は校長先生の調査と同じであるため、

$$D_1 = \overline{X_D} - 1.96 \frac{\sigma}{\sqrt{n}}$$
 , $D_2 = \overline{X_D} + 1.96 \frac{\sigma}{\sqrt{n}}$

となります。

これらに基づいて、どの選択肢が正しいが吟味していきます。

 $C_1 = D_1$ 及び $C_2 = D_2$ になるための必要十分条件は $\overline{X_C} = \overline{X_D}$ が成り立つことですが、 $\overline{X_D}$ が具体的にいくつかわからないため、これが成り立つかどうかは判定できません。 従って①は正しくありません。

$$D_2 - C_1 = \left(\overline{X_D} + 1.96 \frac{\sigma}{\sqrt{n}}\right) - \left(\overline{X_C} - 1.96 \frac{\sigma}{\sqrt{n}}\right) = (\overline{X_D} - \overline{X_C}) + 2 \cdot 1.96 \frac{\sigma}{\sqrt{n}}$$
$$C_2 - D_1 = \left(\overline{X_C} + 1.96 \frac{\sigma}{\sqrt{n}}\right) - \left(\overline{X_D} - 1.96 \frac{\sigma}{\sqrt{n}}\right) = (\overline{X_C} - \overline{X_D}) + 2 \cdot 1.96 \frac{\sigma}{\sqrt{n}}$$

となります。ここに、 $\overline{X_C}=204$ 、 $\sigma=150$ 、n=100を代入すると、

$$D_2 - C_1 = (\overline{X_D} - 204) + 2 \cdot 1.96 \frac{150}{\sqrt{100}} = \overline{X_D} - 145.2$$

$$C_2 - D_1 = (204 - \overline{X_D}) + 2 \cdot 1.96 \frac{150}{\sqrt{100}} = 262.8 - \overline{X_D}$$

となります。 $145.2 < \overline{X_D} < 262.8$ であれば $C_1 < D_2 \ge D_1 < C_2$ が両方成り立つので、一方のみ成り立つとする①は誤りです。

更に、 $\overline{X_D}$ < 145.2であれば D_2 < C_1 となり、 $\overline{X_D}$ < 262.8であれば C_2 < D_1 となるので、 どちらかが成り立つ可能性はあります。つまり②は正しいということです。

ただ、代入して計算するのはメンドイ…(*_*;。試験テクニックとしては、次のようにした方がいいかと思います。まず、①が誤りだとわかっており、③と④と⑤の中で正しいのは一つしかあり得ません。なので、①か②のどちらかは正しいことがわかります。

代入しなくとも、 $\overline{X_C}=\overline{X_D}$ であれば、 $D_2-C_1=C_2-D_1=2\cdot 1.96rac{\sigma}{\sqrt{n}}>0$ だとわかるの

で、片方のみが成り立つとする①は誤りだとわかります。従って、残る②は正しいはずです。このように考えれば、時間を節約することができます。

③、4、⑤に関しては、

$$C_2 - C_1 = \left(\overline{X_C} + 1.96 \frac{\sigma}{\sqrt{n}}\right) - \left(\overline{X_C} - 1.96 \frac{\sigma}{\sqrt{n}}\right) = 2 \cdot 1.96 \frac{\sigma}{\sqrt{n}}$$
$$D_2 - D_1 = \left(\overline{X_D} + 1.96 \frac{\sigma}{\sqrt{n}}\right) - \left(\overline{X_D} - 1.96 \frac{\sigma}{\sqrt{n}}\right) = 2 \cdot 1.96 \frac{\sigma}{\sqrt{n}}$$

となるので、 $C_2 - C_1 = D_2 - D_1$ は必ず成り立ちます。④は正しく、③と⑤は正しくないことがわかります。