2018年6月
問11

一番優しい、統計学の教本

吉田

問 11　次の表は，北海道および沖縄県において，過去 1 年間に野球（キャッチボールを含む）をおこなった 15 歳以上の割合（行動者率）をまとめたものである。

都道府県	標本サイズ	野球の行動者率（%）	15 歳以上の推定人口（千人）
北海道	4,633	7.1	4,542
沖縄県	2,849	9.2	1,150

資料：総務省「平成 28 年社会生活基本調査」

データは単純無作為抽出されたとして，以下の問いに答えよ。ただし，上の表の 15 歳以上の推定人口には誤差がなく真の 15 歳以上人口であるとして答えよ。

〔1〕　北海道における野球の行動者の母比率の 95%信頼区間として，次の ①～⑤ のうちから最も適切なものを一つ選べ。 **22**

① 0.071 ± 0.001　　② 0.071 ± 0.004

③ 0.071 ± 0.007　　④ 0.071 ± 0.010

⑤ 0.071 ± 0.503

〔2〕 次の文章は北海道と沖縄県の全体における野球の行動者の母比率の推定について述べたものである。

「表の数値をそれぞれ

$$n_1 = 4633,\quad \hat{p}_1 = 0.071,\quad N_1 = 4542 \times 10^3,$$
$$n_2 = 2849,\quad \hat{p}_2 = 0.092,\quad N_2 = 1150 \times 10^3$$

とおく。このとき北海道と沖縄県の全体における野球の行動者の母比率の推定値は（ア）となり，その標準誤差は（イ）となる。」

文中の（ア），（イ）にあてはまるものの組合せとして，次の ① 〜 ⑤ のうちから適切なものを一つ選べ。 $\boxed{\textbf{23}}$

① （ア） $\dfrac{N_1\hat{p}_1 + N_2\hat{p}_2}{N_1 + N_2}$ （イ） $\left|\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}} - \sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right|$

② （ア） $\dfrac{N_1\hat{p}_1 + N_2\hat{p}_2}{N_1 + N_2}$

（イ） $\sqrt{\left(\dfrac{N_1}{N_1+N_2}\right)^2 \dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \left(\dfrac{N_2}{N_1+N_2}\right)^2 \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

③ （ア） $\dfrac{\hat{p}_1 + \hat{p}_2}{2}$ （イ） $\dfrac{1}{2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}} + \dfrac{1}{2}\sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

④ （ア） $\dfrac{\hat{p}_1 + \hat{p}_2}{2}$ （イ） $\dfrac{1}{2}$

⑤ （ア） $\hat{p}_1 + \hat{p}_2$ （イ） $\sqrt{\dfrac{1}{n_1 + n_2}}$

知っておくべき知識

■比率の平均値、分散

標準偏差の定義は？
（データが10個の場合）

$$\frac{(\text{データ}1 - \text{平均値})^2 + (\text{データ}2 - \text{平均値})^2 + \cdots + (\text{データ}10 - \text{平均値})^2}{10}$$

をルートしたもの

分散

標準誤差の定義は？
（データが10個の場合）

$$\frac{\text{標準偏差}}{\sqrt{10}}$$

95%信頼区間の定義は？（正規分布の場合）

平均値 ± 1.96 * 標準誤差

この問題のポイント

■比率の**平均値と分散**の定義

比率の平均値: p (割合)

比率の分散: $p(1-p)$

95%信頼区間を求める

$$95\%CI = \text{平均値} \pm 1.96 \times \text{標準誤差}$$

$$= p \pm 1.96 \times \frac{\sqrt{\text{分散}}}{\sqrt{n}}$$

正解：3

$$= 0.071 \pm 1.96 \times \frac{\sqrt{0.071(1-0.071)}}{\sqrt{4633}}$$

$$= 0.071 \pm 0.007$$

〔2〕 次の文章は北海道と沖縄県の全体における野球の行動者の母比率の推定について述べたものである。

「表の数値をそれぞれ

$$n_1 = 4633,\quad \hat{p}_1 = 0.071,\quad N_1 = 4542 \times 10^3,$$
$$n_2 = 2849,\quad \hat{p}_2 = 0.092,\quad N_2 = 1150 \times 10^3$$

とおく。このとき北海道と沖縄県の全体における野球の行動者の母比率の推定値は（ア）となり，その標準誤差は（イ）となる。」

文中の（ア），（イ）にあてはまるものの組合せとして，次の ① 〜 ⑤ のうちから適切なものを一つ選べ。 $\boxed{\mathbf{23}}$

① （ア） $\dfrac{N_1\hat{p}_1 + N_2\hat{p}_2}{N_1 + N_2}$ （イ） $\left|\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}} - \sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right|$

② （ア） $\dfrac{N_1\hat{p}_1 + N_2\hat{p}_2}{N_1 + N_2}$

（イ） $\sqrt{\left(\dfrac{N_1}{N_1+N_2}\right)^2 \dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \left(\dfrac{N_2}{N_1+N_2}\right)^2 \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

③ （ア） $\dfrac{\hat{p}_1 + \hat{p}_2}{2}$ （イ） $\dfrac{1}{2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}} + \dfrac{1}{2}\sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

④ （ア） $\dfrac{\hat{p}_1 + \hat{p}_2}{2}$ （イ） $\dfrac{1}{2}$

⑤ （ア） $\hat{p}_1 + \hat{p}_2$ （イ） $\sqrt{\dfrac{1}{n_1 + n_2}}$

問題を分解しましょう！

〔2〕　次の文章は北海道と沖縄県の全体における野球の行動者の母比率の推定について述べたものである。

「表の数値をそれぞれ

$$n_1 = 4633,\quad \hat{p}_1 = 0.071,\quad N_1 = 4542 \times 10^3,$$
$$n_2 = 2849,\quad \hat{p}_2 = 0.092,\quad N_2 = 1150 \times 10^3$$

とおく。このとき北海道と沖縄県の全体における野球の行動者の母比率の推定値は（ア）となり，その標準誤差は（イ）となる。」

$$\text{母比率の推定値} = \frac{\text{北海道と沖縄の全体の野球行動者}}{\text{北海道と沖縄全体の人口}}$$

$$= \frac{\hat{p}_1 N_1 + \hat{p}_2 N_2}{N_1 + N_2}$$

知っておくべき知識

■平均値、分散、共分散、相関の定義（**暗記！**）

$$E[aX+b]=aE[X]+b$$

$$V[aX+b]=a^2V[X]$$

$$V[X]=E[X^2]-\{E[X]\}^2$$

$$Cov[X,Y]=E[XY]-E[X]E[Y]$$

$$r=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

標準誤差を求めるスタートは、分散を求めること

$$V[aX+b]=a^2V[X]$$

$$V\left[\frac{\hat{p}_1N_1+\hat{p}_2N_2}{N_1+N_2}\right]=V\left[\frac{\hat{p}_1N_1}{N_1+N_2}\right]+V[\frac{\hat{p}_2N_2}{N_1+N_2}]$$

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

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

用語の整理：標準誤差とは？

標準誤差＝推定値の標準偏差

正解：2

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

$$= \left(\frac{N_1}{N_1 + N_2}\right)^2 \times \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \left(\frac{N_2}{N_1 + N_2}\right)^2 \times \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}$$

$$SE = \sqrt{\left(\frac{N_1}{N_1 + N_2}\right)^2 \times \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \left(\frac{N_2}{N_1 + N_2}\right)^2 \times \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$