イチロー選手

Ver.2022.09.03
松浦 真也

疑問 イチロー選手の 2015 年と 2016 年の打率は, それぞれ .229 (398 打数 91 安打) と.291 (327 打数 95 安打) であった. 絶不調だった 2015 年と比べ, 2016 年は復調したと言えるだろうか？
※ 打数＝バッターボックスに立った回数 (厳密には四死球等を除く), 安打＝ヒット

戦略 2015 年も 2016 年も調子が同じだったと仮定. このとき, 打率の差が $0.291 - 0.229 = 0.062$ 以上になる確率を求める (下記の記号を使っていうと, $p_1 = p_2$ のときに, 確率 $P(\overline{p}_2 - \overline{p}_1 \geq 0.062)$ を求める).

Step 1 言葉を定義
母打率: バッターボックスに立った際に, ヒットを打つ確率
標本打率: 年間を通して, 実際にヒットを打った割合, つまり $\dfrac{\text{安打数}}{\text{打数}}$

Step 2 イチロー選手の成績に関し, 記号を定義
<u>2015 年の値</u> p_1: 母打率 (未知), n_1: 打数 (398), h_1: 安打数 (91), $\overline{p}_1 = \dfrac{h_1}{n_1}$: 標本打率 (0.229)
<u>2016 年の値</u> p_2: 母打率 (未知), n_2: 打数 (327), h_2: 安打数 (95), $\overline{p}_2 = \dfrac{h_2}{n_2}$: 標本打率 (0.291)

Step 3 年間の安打数 h_j は, 2 項分布 $Bi(n_j, p_j)$ (2015 年は $j = 1$, 2016 年は $j = 2$) に従うものとすると, 平均 E と分散 V は,

$$E(h_j) = n_j p_j, \qquad V(h_j) = n_j p_j (1 - p_j).$$

これより, $\overline{p}_j = \dfrac{h_j}{n_j}$ の平均, 分散は

$$E(\overline{p}_j) = \frac{1}{n_j} E(h_j) = p_j, \qquad V(\overline{p}_j) = \frac{1}{n_j^2} V(h_j) = \frac{p_j(1 - p_j)}{n_j}.$$

ここで, 打数 n_j が十分大きいとすると, 中心極限定理より, $\overline{p}_j \sim N\left(p_j, \dfrac{p_j(1-p_j)}{n_j}\right)$ と見なせる. よって, 各年の結果が独立に決まるとすると, 正規分布の再生性より,

$$\overline{p}_2 - \overline{p}_1 \sim N\left(p_2 - p_1, \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}\right).$$

Step 4 イチロー選手の調子が2015年と2016年とで, 同じであったと仮定する．つまり, 母打率が等しいとし, $p_1 = p_2 = p$ とおく．このとき,

$$\overline{p}_2 - \overline{p}_1 \sim N\left(0, p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

となる．よって,

$$Z = \frac{\overline{p}_2 - \overline{p}_1}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1).$$

Step 5 イチロー選手の成績のデータから, $n_1 = 398$, $n_2 = 327$, $h_1 = 91$, $h_2 = 95$, $\overline{p}_1 = \dfrac{91}{398}$, $\overline{p}_2 = \dfrac{95}{327}$ である．母打率 p は未知なので, 2年間の通算標本打率で代用．すなわち,

$$p = \frac{2\text{年間の通算安打数}}{2\text{年間の通産打数}} = \frac{h_1 + h_1}{n_1 + n_2} = \frac{91 + 95}{398 + 327} = \frac{186}{725}$$

とする．これらを代入すると,

$$Z = \frac{\frac{95}{327} - \frac{91}{398}}{\sqrt{\frac{186}{725}\left(1 - \frac{186}{725}\right)\left(\frac{1}{398} + \frac{1}{327}\right)}} \approx 1.90.$$

統計数値表 (正規分布表) によれば,

$$P(Z \geq 1.90) \approx 0.028717.$$

結論 イチロー選手の調子が, 2015年も2016年も同じだったにもかかわらず, 打率の差が$0.291-0.229 = 0.062$以上になる可能性は3%程度しかない．この3%という値をどう解釈するかが問題である．例えば, 可能性が5%未満のことは, 滅多に起こらないことだと考えれば, 珍しいことが奇跡的に起こったと考えるよりは, 2015年も2016年も調子が同じだったとする仮定が間違っていると考える方が自然である．従って, 復調した可能性が高いと考えることになる．統計学の言葉で言えば, 「2015年も2016年も調子が同じだった」という帰無仮説を有意水準5%で片側検定すると, 棄却される．

一方, 1%以上の可能性があれば, 十分起こり得ると考えるのであれば, 2015年も2016年も調子が同じだったとしても, おかしくはないと考えることになる（あくまで, 「同じだったとしても, おかしくはない」というだけであり, 「同じだった」と決めつけているわけではない）．統計学の言葉で言えば, 「2015年も2016年も調子が同じだった」という帰無仮説を有意水準1%で片側検定すると, 棄却されない.

補足 Step 5で, 未知の母分散 $p(1-p)$ を標本分散で代用している．このため, 厳密には正規分布ではなく t 分布を用いて確率を求めるべきである．しかし, 本件のように, 標本のサイズ (ここでは打数) が十分大きい場合は, t 分布と正規分布の差は無視できる.

データ出典：MLB Advanced Media （最終閲覧日 2022年9月2日）
`https://www.mlb.com/player/ichiro-suzuki-400085`