ПРОГНОЗ РЕЗУЛЬТАТИВНОСТИ ВЫСОКОКВАЛИФИЦИРОВАННЫХ СПОРТСМЕНОВ

Кутек Т.Б., Ахметов Р.Ф.

Житомирский государственный университет имени Ивана Франко

Аннотация. Ставится и решается основная задача прогноза результативности прыгунов в высоту на базе статистического факторного анализа и экспертного ранжирования полной совокупности антропометрических, технических и специализированных параметров.

Ключевые слова: линейная регрессия, аппроксимация, весовой вектор.

Анотація. Ставиться та вирішується основне завдання прогнозу результативності стрибунів у висоту на базі статистичного факторного аналізу й експертного ранжування повної сукупності антропометричних, технічних і спеціалізованих параметрів.

Ключові слова: лінійна регресія, апроксимація, ваговий вектор.

Annotation. The paper provides ways to solve the major problem of prognosticating the performance of high-jumping based upon statistic and factor analysis as well as expert ranging of the comprehensive unity of anthropometric, technical and specialized parameters.

Key words: linear regression, approximation, weight vector.

Постановка проблемы. В системе спортивной многолетней подготовки большая роль отводится прогнозу результативности как отдельных спортсменов, так и спортивных групп [2; 5]. В связи с этим весьма актуальным является разработка программы прогноза на базе некоторой совокупности информативных параметров спортсменов.

Анализ последних исследований и публикаций. Прогнозирование основывается на использовании метода экстраполяции, предполагающего распространение выводов, полученных из наблюдения над одной частью какого-либо явления, на другие его части [2; 6; 7]. В условиях спорта экстраполяция позволяет осуществить прогнозы роста результативности на основе изучения соответствующих закономерностей в предшествующие годы. Задачу прогноза результативности спортсменов можно решить на базе факторного анализа и динамики развития физических параметров и результатов на некотором ограниченном интервале времени [1; 6]. Особое внимание уделяется «раннему» прогнозу на период до 17 лет по данным начального периода (10-13 лет).

Цель работы - разработка программы прогнозирования результативности высококвалифицированных спортсменов.

Результаты исследования. Поскольку результаты и физические параметры спортсменов в группе имеют случайный разброс (дисперсию), то, говоря о задаче прогноза результативности, имеет смысл рассматривать прогноз средней результативности $\overline{\mathrm{H}}(t)$, как функции средних по группе физических параметров $\vec{X}_P$, которые будем представлять в виде матрицы столбца:

$$\vec{\mathrm{X}}_P = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_P \end{pmatrix}, \text{ P=1,2,…,N-1; N} \geq 2,$$

где N – полное число спортивных параметров, включая сам результат (H). Полное множество P-мерных группировок из (N-1) по P равно числу сочетаний из (N-1) по P :

$$\vec{\mathrm{X}}_P \in U_{\vec{X}_P} = \{ \vec{X}_P^{\alpha}, \alpha = 1,2,...,C_{N-1}^P \}, \qquad (1)$$

$$C_{N-1}^P = \frac{(N-1)!}{P!(N-1-P)!}.$$

Информативность различных P-мерных группировок $\vec{\mathrm{X}}_P$в задачах прогноза результативности будет также различной. Вопрос о выборе оптимальной совокупности наиболее информативных параметров из множества (1) при различных P требует самостоятельных глубоких исследований в рамках отдельной НИР. В данной работе предлагается один из альтернативных вариантов решения задачи, который вполне приемлем с точки зрения точности прогноза. В первом приближении рассматривается задача линейного прогноза в рамках классической теории линейной регрессии (интерполяции) в математической статистике [4; 7-8]. Речь идет о нахождении аппроксимации

$$\overline{\mathrm{H}} \cong \mathrm{H}_0 + \alpha_1 X_1 + \alpha_2 X_2 + ... + \alpha_P X_P, \qquad (2)$$

где $\mathrm{H}_0, \alpha_1, ..., \alpha_P$ – неизвестные параметры регрессии, которые требуется оценить по данным некоторого количества возрастных групп. В более точной постановке приближенная линейная регрессия (2) представляется в виде:

$$\overline{\mathrm{H}}(t) = \mathrm{H}_0 + \alpha_1 X_1(t) + \alpha_2 X_2(t) + ... + \alpha_P X_P(t) + \xi(t), t \in T = (a,b), \qquad (3)$$

где $\xi(t)$ – ошибка прогноза с нулевым средним ($M\xi(t) = 0$) и неизвестной дисперсией $\sigma_\xi^2 = M\xi^2$ (M – оператор математического ожидания – среднего). Если в результате решения задачи линейной регрессии на интервале времени T получены оценки неизвестных параметров регрессии:

$$\mathrm{H}_0 = \hat{\mathrm{H}}_0(T); \; \alpha_n = \hat{\alpha}_n(T), \quad n = 1,2,...,P,$$

то прогнозное значение средней результативности вне этого интервала представляется в виде:

$$\overline{\mathrm{H}}^{\wedge}(t_0) = \hat{\mathrm{H}}_0(T) + \sum_{n=1}^{P} \hat{\alpha}_n(T) X_n(t_0), \quad t_0 > b, \qquad (4)$$

где набор физических параметров $\{ X_n(t_0), \quad n = 1,2,...,P \}$ – задается на прогнозируемый момент времени t_0. При этом среднеквадратическая ошибка (СКО) прогноза оценивается величиной $\sigma_\xi(T)$. Насколько «удачно» получена оценка (4), – зависит от многих факторов и последнее слово здесь за практикой (экспериментальной апробации). Проведенная в данной работе апробация модели (4) показывает, что она практически вполне приемлема. СКО (среднее квадратическое отклонение) при этом не превышает 3-х сантиметров, а прогнозируемый рекордный результат составляет 250 см.

Матричное решение задачи линейной регрессии результативности по заданной совокупности наиболее информативных параметров

Для оценки параметров регрессии $\mathrm{H}_0, \alpha_1, ..., \alpha_P$ составляется следующая система линейных алгебраических уравнений:

$$\mathrm{H}_0 + \sum_{m=1}^{P} \alpha_m X_m(t_1) = \overline{\mathrm{H}}(t_1)$$

$$\mathrm{H}_0 + \sum_{m=1}^{P} \alpha_m X_m(t_2) = \overline{\mathrm{H}}(t_2) \qquad (5)$$

$$\mathrm{H}_0 + \sum_{m=1}^{P} \alpha_m X_m(t_N) = \overline{\mathrm{H}}(t_N),$$

где в данном разделе, следуя стандартным обозначениям, N – число возрастных групп (в данной работе N<9). Система (5) представляется в матричном виде:

$$\mathrm{H}_0 \vec{1}_N + \sum_{m=1}^{P} \alpha_m \vec{X}_N^m = \vec{\overline{\mathrm{H}}}_N \Rightarrow \qquad (6)$$

$$\vec{1}_N = \begin{pmatrix} 1 \\ 1 \\ .. \\ 1 \end{pmatrix}_N, \quad \vec{X}_N{}^m = \begin{pmatrix} X_m(t_1) \\ X_m(t_2) \\ .... \\ X_m(t_N) \end{pmatrix}, \quad \vec{\overline{\mathrm{H}}}_N = \begin{pmatrix} \overline{\mathrm{H}}(t_1) \\ \overline{\mathrm{H}}(t_2) \\ .... \\ \overline{\mathrm{H}}(t_N) \end{pmatrix}.$$

Вводя т.н. «сигнальный» регрессионный вектор (СРВ):

$$\vec{s}_M = \begin{pmatrix} \mathrm{H}_0 \\ \alpha_1 \\ ... \\ \alpha_P \end{pmatrix}_M = \begin{pmatrix} s_1 \\ s_2 \\ ... \\ s_M \end{pmatrix}, \quad M = P + 1, \qquad (7)$$

$$s_1 = \mathrm{H}_0, \; s_2 = \alpha_1, \; s_3 = \alpha_2, ..., s_M = \alpha_P,$$

матричную систему (6) представляем также в стандартном виде:

$$\sum_{m=1}^{M} s_m \vec{Y}_N^m = \vec{\overline{\mathrm{H}}}_N \quad \Rightarrow \quad Y_{NM} \vec{s}_M = \vec{\overline{\mathrm{H}}}_N, \qquad (8)$$

$$\vec{Y}_N^1 = \vec{1}_N, \vec{Y}_N^2 = \vec{X}_N^1, ..., \quad \vec{Y}_N^M = \vec{X}_N^P, \quad Y_{NM} = (\vec{Y}_N^1 \vec{Y}_N^2 ... \vec{Y}_N^P),$$

где Y_{NM} – измеримая матрица наблюдений (ИМН); $\vec{\overline{\mathrm{H}}}_N$ – измеримый вектор средних результатов (ВСР).

Согласно общей теории линейной регрессии система (8) может быть решена, если она полностью определена или переопределена:

$$N \geq M + 1 = P + 2 \Rightarrow \quad Rank Y_{NM} = M. \qquad (9)$$

Отметим, что величина (M+1) обусловлена тем, что в число неизвестных помимо M=P+1 неизвестных параметров регрессии необходимо включить также и неизвестное СКО σ_ξ. При выполнения условия (9) статистическое решение задачи линейной регрессии представляется в виде:

$$\vec{s}_M^{\wedge} = Y_{NM}^{-} \vec{\overline{\mathrm{H}}}_N, \; Y_{NM}^{-} = (Y_{NM}^T Y_{NM})^{-1} Y_{NM}^T, \qquad (10)$$

$$(\sigma_\xi^2)^{\wedge} = \frac{1}{N-M} // \vec{\overline{\mathrm{H}}}_N^{\wedge} - \vec{\overline{\mathrm{H}}} //^2 = \frac{// \Lambda_{NN}^{M\perp} \vec{\overline{\mathrm{H}}}_N //^2}{N-M}, \qquad (11)$$

$\vec{\vec{H}}_N^{\wedge} = Y_{NM}\vec{s}_M^{\,\wedge} = \Lambda_{NN}^M, \quad \Lambda_{NN}^M = Y_{NM}Y_{NM}^-, \quad \Lambda_{NN}^{M\perp} = I_{NN} - \Lambda_{NN}^M$,

$Rank\Lambda_{NN}^M = M, \quad Rank\Lambda_{NN}^{M\perp} = N - M$,

где Y_{NM}^- – псевдообратная матрица [5]; Λ_{NN}^M – проектор в линейную оболочку из базисных векторов $\{\vec{Y}_N^m, m = 1,2,...,M\}$; $\Lambda_{NN}^{M\perp}$ – ортогональный вектор.

В данной работе наиболее точное решение получено в случае P=3 при различных N с учетом необходимого условия разрешения (9):

$$5 \le N \le 8. \qquad (12)$$

Разработана специализированная программа corrS1m.com в среде Turbo Pascal. Результаты расчетов приведены в приложении. Специфической математической особенностью задачи регрессии спортивного результата является то, что в силу довольно однородного состава групп столбцовые вектора ИМН Y_{NM} оказываются хотя и случайными, но с малым угловым расхождением относительно «единичного» вектора $\vec{1}_N$. Последнее обстоятельство требует жесткого контроля точности обращения матрицы Грама $(Y_{NM}^T Y_{NM})_{MM}$, т.к. в случае высокой угловой корреляции («схожести») векторов $\vec{Y}_N^m$ матрица Грама оказывается часто плохо обусловленной [3] с большим динамическим диапазоном собственных чисел в области малых величин. При этом точность обращения матрицы Грама с ростом размерности P>3 (числа учитываемых информативных параметров) начинает резко падать и дальнейшее увеличение размерности P не представляется возможным.

Отметим также, что в данной работе максимальное число возрастных групп матеріальної крапки 8.

Поэтому в силу условия (9) предельное число наиболее информативных параметров ограничивается величиной 6:

$$P \le N - 2 \le N_{\max} - 2 = 8 - 2 = 6. \qquad (13)$$

Как уже отмечалось, размерность вектора информативных физических параметров спортсменов $\vec{X}_P$ принципиально ограничивается числом возрастных групп и в данной работе оказывается не более шести (13). Прямой перебор различных возможных комбинаций информативных параметров из числа сочетаний (1) оказывается достаточно большим и весьма трудоемким. Так для p=3 и N=8 требуется перебрать

$$C_5^3 + C_6^3 + C_7^3 + C_8^3 = 10 + 15 + 21 + 15 = 102 \qquad (14)$$

комбинации 3-х мерных совокупностей $\vec{X}_P$.

Апробация алгоритмов прогноза результативности прыгунов в высоту по различному числу информативных параметров

Программа РЕГРЕССИЯ (corrS1m.com) содержит следующие разделы:

1. Вызов исходных статистических данных (файл g1_21_9.dat)
2. Шифр файла: tN-M, где N – число возрастных групп, по которым проводится прогноз на будущее; M – число информативных параметров (N$\ge$ M+2)
3. Выбор M информативных параметров (из номеров 2-21).
4. Анализ ранга регрессионной матрицы $Y_{N(M+1)}$ методом Грама-Шмидта.

5. Анализ корреляции информативных параметров по годам.

6. Спектральный анализ матрицы Грама $Y^T Y$ размером (M+1)*(M+1).

7. Оценка точности обращения матрицы Грама.

8. Оценка статистических характеристик информативных параметров (средние, СКО, корреляционная матрица).

9. Решение задачи линейной регрессии.

10. Оценка дисперсии шума (СКО=s).

11. Прогнозирование за пределы выбранных возрастных групп, включая прогноз рекордных результатов (приводятся соответствующие графики).

Результаты расчетов приводятся в приложении.

Выводы

Задача прогноза результативности спортсменов является задачей интерполяции средней (по возрастной группе) результативности ($\overline{\text{H}}$) в виде линейной комбинации средних значений наиболее информативных физических параметров спортсменов $(\bar{x}_1, \bar{x}_2, ..., \bar{x}_P)$ с указанием точности (СКО) прогноза:

$$\overline{\text{H}} = \text{H}_0 + \alpha_1 \bar{x}_1 + \alpha_2 \bar{x}_2 + ... + \alpha_P \bar{x}_P + \xi, \quad \overline{\xi^2} = \sigma_\xi^2,$$

где $\text{H}_0, \alpha_1, \alpha_2, ..., \alpha_P$ – параметры регрессии; σ_ξ – СКО прогноза.

В условиях априорной неопределенности об СКО прогноза необходимым условием решения задачи прогноза является превышение числа используемых возрастных групп ($N_{\text{ВГ}}$) над числом используемых информативных физических параметров (P), как минимум на две единицы:

$N_{\text{ВГ}} \geq P+2$.

Так, при числе информативных физических параметров P=3 требуются средние значения по 5-ти возрастным группам (10, 11, 12, 13 и 14 лет). При этом можно дать прогноз результативности не только на любой «внутренний» момент времени t_0 ($10 \leq t_0 \leq 13$), но и на будущие моменты времени t_0>13, включая прогноз рекордных результатов. Для этого достаточно в полученную формулу регрессии подставить значения прогнозных средних значений физических параметров $\{ \bar{x}_n(t_0), n = 1,2,...,P \}$:

$$\overline{\text{H}}(t_0) \cong \text{H}_0 + \alpha_1 \bar{x}_1(t_0) + \alpha_2 \bar{x}_2(t_0) + ... + \alpha_P \bar{x}_P(t_0) \quad (\pm\sigma_\xi).$$

В частности, при прогнозе по трем параметрам (x_{12},x_9,x_{21}) по 5-ти возрастным группам (10, 11, 12, 13 и 14 лет) получена следующая регрессионная функция:

$$\hat{H} = 0.514 + 0.715 x_{12} + 0.031 x_9 + 0.819 x_{21}, \quad \hat{\sigma} = s = 0.9 \tilde{n}\grave{i} \ .$$

где x_{12} – высота вылета ОЦТ; x_9 – скорость вылета ОЦТ; x_{21} – прыжок вверх с трёх шагов. При этом прогнозное значение результата для мастеров спорта международного класса составляет 2,36 см, что отличается от их среднего результата (2.33 см) всего на 3 см.

Литература:

1. Ахметов Р.Ф. Теоретико-методичні основи управління системою багаторічної підготовки спортсменів швидкісно-силових видів спорту: автореф. дис. на здобуття наук, ступеня д-ра наук з фіз. виховання і спорту / Р.Ф. Ахметов.

– К., 2006. – 39 с.
2. Баландин В.И., Блудов Ю.М., Плахтиенко В.А. Прогнозирование в спорте. – М.: Физкультура и спорт, 1986. – 193 с.
3. Гантмахер Ф.Р. Теория матриц. – 4-е изд.– М.: Наука, 1988. – 552 с.
4. Крамер Г. Математические методы статистики: Пер. с англ. / Под ред. академика А.Н. Колмогорова. – М.: Мир, 1975. – 648 с.
5. Платонов В.Н. Система подготовки спортсменов в олимпийском спорте. Общая теория и ее практические приложения: учеб. для студентов вузов физ. воспитания и спорта: утв. М-вом образования и науки Украины / В.Н. Платонов. – К.: Олимп. л-ра, 2004. – 807 с.
6. Плахтиенко В.А., Мельник В.Г. Прогнозирование в спорте. – Л.: ВДКИФК, 1980. – 79 с.
7. Пугачев В.С. Теория вероятностей и математическая статистика. – М.: Наука, 1979. – 496 с.
8. Harman H.H. Modern factor analysis. – University of Chicago Press, 1960. Русский перевод: Современный факторный анализ. – М.: Статистика, 1972. – 516 с.
9. Lawley D.N., Maxwell A.E. Factor analysis as a statistical method. – Butterworths. – London, 1963. Русский перевод: Факторный анализ как статистический метод. – М.: Мир, 1967. – 413 с.

Информация об авторах:
Кутек тамара Борисовна, доктор наук по физическому воспитанию и спорту, декан факультета физического воспитания, Житомирский государственный университет имени Ивана Франко
Ахметов Рустам Фагимович, доктор наук по физическому воспитанию и спорту, профессор, Житомирский государственный университет имени Ивана Франко

Поступила в редакцию 20.03.2015