ПРИМЕНЕНИЕ МЕТОДА ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ В ИССЛЕДОВАНИЯХ В СФЕРЕ ФИЗИЧЕСКОЙ КУЛЬТУРЫ И СПОРТА

Нечипорук Ю., Шутенко Е.
Научный руководитель А.А.Хармонова
Смоленская государственная академия физической культуры, спорта и туризма

С каждым годом методы математической статистики все шире используются в сфере физической культуры и спорта, они позволяют оценить перспективность спортсменов, условия, наиболее благоприятные для тренировок, их эффективность, с их помощью обрабатываются показания датчиков, контролирующих нагрузки спортсменов и т.д. Любое проводимое исследование в области физической культуры и спорта предполагает не только организацию сбора, стандартной записи, систематизации и обработки экспериментальных статистических данных, но также их интерпретацию и получение научных и практических выводов, что невозможно без применения методов математической статистики.

Поскольку большинство исследований проводятся с использованием выборочных данных, то проекция результатов на генеральную совокупность всегда содержит элемент неточности выборочной точечной оценки исследуемого параметра, а при незначительном числе статистических данных, точечная оценка может существенно отличаться от оцениваемого параметра и приводить зачастую к значительным ошибкам. В таких случаях следует пользоваться интервальными оценками. Наиболее распространенным методом интервального оценивания является метод доверительных интервалов. Смысл вычисления доверительного интервала заключается в построении по данным выборочной совокупности такого интервала, чтобы можно было утверждать с заданной вероятностью, что значение оцениваемого параметра находится в этом интервале.

В данной работе мы рассмотрим несколько примеров применения метода доверительного интервала в исследованиях в области физической культуры и спорта.

1. *Определение границ для среднего значения исследуемого параметра генеральной совокупности.*

По известной величине выборочной характеристики можно определить интервал, в котором с той или иной вероятностью определяется значение параметра генеральной совокупности, оцениваемого по этой выборочной характеристике. Вероятности, признанные достаточными для того, чтобы уверенно судить о генеральных параметрах на основании выборочных характеристик, называются доверительными. Обычно в качестве доверительных вероятностей выбирают значения 0,9; 0,95; 0,99 и 0,999.

Интервал, в котором с заданной доверительной вероятностью находится оцениваемый генеральный параметр, называется доверительным интервалом.

Доверительный интервал для выборок малого объема $(n < 30)$ вычисляется по формуле $\left(\bar{x} - t_\alpha \cdot \frac{s}{\sqrt{n}}; \bar{x} + t_\alpha \cdot \frac{s}{\sqrt{n}}\right)$.

Доверительный интервал для среднего значения параметра генеральной совокупности при $n \geq 30$ записывается в следующем виде $\left(\bar{x} - u_\alpha \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + u_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$.

Среднее значение исследуемого параметра генеральной совокупности является хотя и неизвестным, но фиксированным параметром, а границы доверительного интервала, полученные по случайной выборке n, будут также случайными величинами. Когда доверительная вероятность выбирается 95%, это означает, что примерно в 95% случаев фиксированное, но неизвестное среднее значение исследуемого параметра генеральной совокупности окажется в границах доверительного интервала.

Рассмотрим конкретный пример. Средний рост в выборке баскетболистов $(n = 225)$ составил 184,65 см, среднее квадратическое отклонение равно 0,65 см. Составить доверительный интервал для среднего значения роста баскетболистов с доверительной вероятностью 0,95.

Объем выборки $n = 225$, т.е. для определения доверительного интервала необходимо воспользоваться формулой для большого объема выборки $\left(\bar{x} - u_\alpha \cdot \frac{\sigma}{\sqrt{n}} ; \bar{x} + u_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$. Среднее арифметическое и среднее квадратическое отклонение уже вычислены: $\bar{x} = 184{,}65$, $\sigma = 0{,}65$. При доверительной вероятности 95% по таблице определяем $u_\alpha = 1{,}96$.

Определяем границы доверительного интервала:
$$\left(184{,}65 - 1{,}96 \cdot \frac{0{,}65}{\sqrt{225}} ; 184{,}65 + 1{,}96 \cdot \frac{0{,}65}{\sqrt{225}}\right) = (184{,}56 ; 184{,}74).$$

Таким образом, с вероятностью 95% можем утверждать, что истинное значение среднего роста баскетболистов находится в интервале (184,56; 184,74) см.

2. *Прогнозирование спортивных результатов отдельных спортсменов.*

Рассмотрим задачу прогнозирования результата спортсмена на соревнованиях при помощи результатов, показанных им в процессе подготовки. При этом учитываем, что любое предсказанное значение какого-либо параметра, вычисленное на основе ограниченного числа опытов, всегда будет содержать элемент случайности, поэтому это значение называют оценкой соответствующего параметра.

Вероятность попадания прогнозируемого результата в интервал $(\bar{x} - \varepsilon ; \bar{x} + \varepsilon)$ равна α. Доверительную вероятность задают заранее, полагая

α равной числу, близкому к единице, $\bar{x}$ – среднее арифметическое полученных в опыте значений исследуемого параметра, ε - точность оценки.

Для заданной вероятности α величина ε находится по формуле $\varepsilon = \sqrt{\frac{D}{n}} \cdot \arg\Phi\left(\frac{1+\alpha}{2}\right)$, где $\arg\Phi\left(\frac{1+\alpha}{2}\right)$ является функцией, обратной функции Лапласа, т.е. $\arg\Phi\left(\frac{1+\alpha}{2}\right)$ является таким значением аргумента, при котором значение функции Лапласа равно $\frac{1+\alpha}{2}$.

Рассмотрим конкретный пример. Спринтер в течение последнего месяца подготовки к соревнованиям 20 раз пробегал стометровку и показал следующие результаты: 10,5; 10,8; 11,2; 10,9; 10,4; 10,6; 10,9; 11,0; 10,3; 10,8; 10,6; 11,3; 10,5; 10,7; 10,8; 10,9; 10,8; 10,7; 10,9; 11,0. Требуется оценить ожидаемый на соревнованиях результат и найти доверительный интервал для доверительной вероятности $\alpha = 0{,}9$.

Произведя необходимые вычисления находим среднее значение $\bar{x} = 10{,}78$ и несмещенную дисперсию $D = 0{,}064$. По таблице находим $\arg\Phi\left(\frac{1+0{,}9}{2}\right) = 1{,}643$.

Таким образом, $\varepsilon = \sqrt{\frac{0{,}064}{20}} \cdot 1{,}643 = 0{,}093$.

Границы доверительного интервала при этом оказываются равными $(10{,}78 - 0{,}093\,; 10{,}78 + 0{,}093) = (10{,}69\,; 10{,}87\,)$. Таким образом, с вероятностью 90% спортсмен покажет результат между 10,69 и 10,87.

3. Определение необходимого числа исследуемых спортсменов.

При подготовке к проведению исследования достаточно часто исследователя интересует вопрос: какой минимальный объем выборки необходим для того, чтобы полученный средний результат (точечная оценка) исследуемого признака находился в определенных пределах от истинного среднего значения данного признака генеральной совокупности? Ответить на этот вопрос можно, если ввести доверительную вероятность и выбрать объем выборки n таким образом, чтобы доверительный интервал имел заданный размер.

Требуется, чтобы выборочное среднее $\bar{x}$ отличалось от генерального $\bar{X}$ не более чем на заданную величину ε. Это означает, что половина ширины доверительного интервала должна быть равна ε, т.е. половина от

$$\left(\bar{x}+u_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)-\left(\bar{x}-u_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)=2u_\alpha \frac{\sigma}{\sqrt{n}}$$

должна равняться ε:

$$u_\alpha \frac{\sigma}{\sqrt{n}}=\varepsilon.$$

Отсюда требуемый объем выборки определяется следующим образом:

$$n=\left(\frac{u_\alpha \cdot \sigma}{\varepsilon}\right)^2,$$

где α - доверительная вероятность, u_α - критическое значение стандартизованного нормального распределения (определяется на основании таблицы), ε - точность оценки (приемлемая ошибка выборочного исследования), σ - среднее квадратическое отклонение.

Как правило, доверительная вероятность равна 95%. Если требуется увеличить доверительную вероятность, то выбирают величину равную 99%. Если можно ограничиться более низкой вероятностью, выбирают 90%.

Для определения объема используется величина u_α, а не t, поскольку для вычисления критического значения t размер выборки необходимо знать заранее.

Как правило, среднее квадратическое отклонение никогда неизвестно, и в большинстве случаев его оценивают на основе предшествующих исследований. Также можно выполнить пилотный проект и вычислить среднее квадратическое отклонение по результатам.

Рассмотрим конкретный пример. Цель исследования состояла в определении среднего времени стартовой реакции спринтеров высокой квалификации. На основании ранее полученных данных установлено, что среднее квадратическое отклонение этого показателя составляет 0,07 сек. Какой минимальный объем выборки (минимальное число спортсменов, участвующих в обследовании) необходим для того, чтобы среднее значение, определенное по выборке, отличалось от истинного значения изучаемого показателя не более чем на 0,01 сек?

Выберем доверительную вероятность равную 95%, тогда $u_\alpha = 1{,}96$, точность $\varepsilon = 0{,}01$, среднее квадратическое отклонение $\sigma = 0{,}07$. Все данные подставим в формулу, для определения минимального объема выборки:

$$n = \left(\frac{u_\alpha \cdot \sigma}{\varepsilon}\right)^2 = \left(\frac{1{,}96 \cdot 0{,}07}{0{,}01}\right)^2 = 188.$$

Таким образом, при объеме выборки $n = 188$ существует 95%-ная вероятность того, что выборочное среднее арифметическое будет отличаться от генерального среднего не более чем на 0,01сек.