

Ямнов Е.С.

МАШИННОЕ ОБУЧЕНИЕ ЛИНЕЙНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ ПАРАМЕТРОВ АНТРОПОМЕТРИИ ДЕВУШЕК АСТЕНИЧЕСКОГО СОМАТОТИПА

Научный руководитель: д-р мед. наук, проф. Волошин В.Н.

Кафедра топографической анатомии и оперативной хирургии

Луганский государственный медицинский университет имени Святителя Луки, г. Луганск

Актуальность. Регрессионные модели в машинном обучении являются алгоритмами, которые применяются для прогнозирования значений переменной отклика в ответ на изменение значений переменных-предикторов. В настоящее время в публикациях, посвященных антропометрическим исследованиям, крайне редко можно встретить сведения о применении методов машинного обучения. Отсутствие таких данных в отношении параметров антропометрии жителей Луганской Народной Республики обуславливает актуальность настоящего исследования.

Цель: провести машинное обучение регрессионных моделей некоторых параметров антропометрии девушек астеников.

Материалы и методы. В исследовании приняли участие 117 девушек астеников, проживающих в Луганской Народной Республике. Все девушки дали письменное информированное согласие на проведение исследования. У девушек общепринятыми методами определяли длину тела стоя (ДТ) с точностью до 0,5 см и массу тела (МТ) с точностью до 0,1 кг. В качестве зависимой переменной определена ДТ, а в качестве предиктора – МТ. Материал для представленной работы собран с соблюдением правил биоэтики и, согласно закону о защите личных данных, при дальнейшей обработке был деперсонифицирован. Статистическая обработка полученных данных проведена в среде программирования R с применением пакетов readxl, stats, tidyverse, car, DescTools, DAAG. Регрессионный анализ указанных выше переменных осуществляли методом наименьших квадратов. Независимость остатков модели определяли с помощью теста Wald-Wolfowitz. Аутокорреляцию остатков проверяли путем выполнения теста Durbin-Watson. Из первичной модели (ПМ) удаляли «влияющие» значения и получали новую модель (НМ), на основе которой формировали модель для обучения. Проводили 3-кратное обучение полученной модели. Прогностическую способность моделей оценивали с помощью информационного критерия Akaike (AIC). Критический уровень значимости полученных в работе статистических критериев определен на уровне 0,05.

Результаты и их обсуждение. Коэффициенты линейной регрессии ДТ~МТ β_0 и β_1 ПМ определены на уровне соответственно 128,94 ($p<0,001$) и 0,72 ($p<0,001$). Доверительные интервалы последних находятся в пределах [120,70; 137,20] и [0,56; 0,87] соответственно. Таким образом, формула линейной регрессии ПМ принимает вид $ДТ(см)=128,94+0,72*МТ(кг)$. F-статистика данной модели определена на уровне 84,74 ($p<0,001$). Приведенные выше результаты указывают на статистическую значимость ПМ. После удаления из первичного набора данных «влияющих» значений в количестве шести, получен набор данных, на основе которого построена НМ. AIC последней определен на уровне 725,393. Сравнение ПМ и НМ указывает на статистически значимое увеличение прогностических свойств НМ ($F=99,10$; $p<0,001$). Формула регрессии НМ имеет вид $ДТ(см)=122,51+0,83*МТ(кг)$. Обучение НМ позволило установить значения среднеквадратических ошибок на уровне 40,09 (первая итерация), 19,38 (вторая итерация) и 28,08 (третья итерация).

Выводы. Изменение значения зависимой переменной (ДТ) девушек астенического соматотипа при изменении значения предиктора (МТ) описывается формулой $ДТ(см)=122,51+0,83*МТ(кг)$. Качество обучения полученной модели подтверждается значениями метрик регрессионного анализа.