МИНИСТЕРСТВО СПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский университет спорта «ГЦОЛИФК»

**Кафедра теории и методики единоборств**

Ермаков А.В.

**ПОСТРОЕНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ ДЛЯ ИССЛЕДОВАНИЯ СПОРТИВНОЙ ДЕЯТЕЛЬНОСТИ В ЕДИНОБОРСТВАХ**

для обучающихся

по направлению подготовки 49.03.04 Спорт

г. Москва, 2024 г.

Утвержден и рекомендован

к использованию в образовательном процессе

Экспертно-методическим советом

Социально-гуманитарного факультета Протокол № от « » 20г.

Ермаков А.В.

Построение регрессионных моделей для исследования спортивной деятельности в единоборствах: учебное пособие для обучающихся по направлению подготовки 49.03.04 Спорт. - г. Москва: РУС «ГЦОЛИФК», 2024. - 48 с.

Рецензент: Моисеев Александр Борисович – доцент кафедры Теории и методики фехтования, современного пятиборья и стрелковых видов спорта РУС «ГЦОЛИФК», кандидат педагогических наук, доцент.

Учебное пособие посвящено построению регрессионных моделей для исследования спортивной деятельности в единоборствах и адресован обучающимся по программам бакалавриата для подготовки к практическим занятиям и к зачету по дисциплине Теория и методика единоборств

Содержание

Оглавление

[Введение 4](#_Toc187338666)

[Применение интерполяции, аппроксимации и экстраполяции в изучении спортивной деятельности 10](#_Toc187338667)

[Теоретико-методические основы регрессионного анализа спортивной деятельности 24](#_Toc187338668)

[Использование регрессионного анализа для проведения исследований спортивной подготовки (на примере технико-тактической подготовки). 39](#_Toc187338669)

[Сравнительный анализ соревновательной деятельности с помощью регрессионных моделей 50](#_Toc187338670)

[Список литературы 63](#_Toc187338671)

# Введение

Целью данного учебного пособия является снабжение студентов бакалавриата, обучающихся по направлению подготовки 49.03.04 Спорт теоретическими знаниями и практическими умениями построения регрессионных моделей для исследования спортивной деятельности в единоборствах. В результате применения пособия будущие бакалавры обучаются эффективно воздействовать и управлять ходом подготовки спортсменов различных гендерных и возрастных групп, а также уровня квалификации. Обучаемые по программе бакалавриата по дисциплине Теория и методика единоборств получают набор знаний и умений по следующим компетенциям (табл.1)

Таблица 1. – Компетенции реализуемые при построении регрессионных моделей

|  |  |
| --- | --- |
| ОПК-1 | Способен планировать содержание занятий физической культурой и спортом в рамках сферы спортивной подготовки, сферы образования с учетом положений теории и методики физической культуры, теории спорта, анатомо-морфологических, физиологических и психических особенностей занимающихся различного пола и возраста. |
| ОПК-1.1. | Знает положения теории физической культуры теории спорта, физиологические, анатомо-морфологические и психические особенности занимающихся физической культурой и спортом различного пола и возраста; специфику, масштабы и предметные аспекты планирования, его объективные и субъективные предпосылки в рамках сферы спортивной подготовки и сферы образования; технологии разработки программ спортивной подготовки и образовательных программ в области физической культуры и спорта. |
| ОПК-1.2. | Умеет планировать тренировочный процесс на этапах спортивной подготовки и содержание занятий физической культурой и спортом в рамках спортивной подготовки и сферы образования; определять методы, формы и средства при разработке программ спортивной подготовки и образовательных программ по физическому воспитанию обучающихся. |
| ОПК-3.2. | Умеет подбирать средства, методы и приемы различных видов физкультурно-спортивной деятельности для проведения занятий, составлять комплексы упражнений; проводить занятия физической культурой и спортом в сфере спортивной подготовки и сфере образования. |
| ОПК-4.2. | Умеет подбирать средства и методы развития физических качеств, повышения функциональных возможностей спортсменов и обучающихся в соответствии со спецификой вида физкультурно-спортивной деятельности; подбирать средства, методы и приемы психолого-педагогического сопровождения в сфере спортивной подготовки и сфере образования. |
| ОПК-5.2. | Умеет подбирать и использовать средства и методы обучения и совершенствования техники выполнения упражнений; составлять программы тренировочных занятий различной направленности в спортивной подготовке со спортсменами различной квалификации и различным контингентом обучающихся; организовывать деятельность спортсменов и обучающихся в процессе реализации программ спортивной подготовки и образовательных программ в области физической культуры и спорта; обеспечивать участие спортсменов и обучающихся в спортивных и физкультурных мероприятиях; выбирать методы индивидуализации, осуществлять процесс подготовки с учетом индивидуальных особенностей спортсменов и обучающихся; применять инновационные методики физической культуры и спорта в профессиональной деятельности. |
| ОПК-9 | Способен анализировать соревновательную деятельность для корректировки педагогического воздействия на спортсменов и обучающихся. |
| ОПК-9.2. | Умеет анализировать соревновательную деятельность для корректировки процесса подготовки спортсменов и обучающихся. |
| ОПК-12 | Способен осуществлять контроль технической, физической, тактической, психологической, интеллектуальной и интегральной подготовленности спортсменов, физического развития спортсменов и обучающихся, в том числе с использованием методик измерения и оценки. |
| ОПК-12.1. | Знает содержание и формы, методы контроля, методики измерения и оценки физического развития и функционального состояния организма занимающихся физической культурой и спортом; технической, физической, тактической, психологической, интеллектуальной и интегральной подготовленности спортсменов и обучающихся; психического состояния спортсменов и обучающихся. |
| ОПК-12.2. | Умеет подбирать методики измерения и оценки физического развития, подготовленности, психического состояния спортсменов и обучающихся; осуществлять контрольные процедуры; интерпретировать результаты контроля физического развития, подготовленности, психического состояния спортсменов и обучающихся. |
| ОПК-13 | Способен использовать результаты педагогического, психологического и медико-биологического контроля для коррекции тренировочного процесса в избранном виде спорта, осуществлять контроль за формированием общей культуры, воспитания личностных качеств у лиц, занимающихся физической культурой и спортом. |
| ОПК-13.1. | Знает особенности педагогического, психологического и медико-биологического контроля в процессе спортивной подготовки в избранном виде спорта; алгоритмы подготовки рекомендаций по коррекции тренировочного процесса в избранном виде спорта; основы формирования общей культуры, воспитания личностных качеств у лиц, занимающихся физической культурой и спортом. |
| ОПК-13.2. | Умеет разрабатывать рекомендации для коррекции тренировочного процесса на основе анализа результатов комплексного контроля в избранном виде спорта; определять уровень общей культуры и личностные качества у лиц, занимающихся физической культурой и спортом. |
| ОПК-15 | Способен проводить научные исследования по определению эффективности используемых средств и методов в сфере спортивной подготовки и сфере образования. |
| ОПК-15.1. | Знает основы научно-методической деятельности, научную терминологию, принципы, средства, методы и технологию организации научного исследования в сфере спортивной подготовки и сфере образования. |
| ОПК-15.2. | Умеет разрабатывать и реализовывать программу научного исследования по определению эффективности используемых средств и методов физической культуры и спорта. |

Цифровые технологии сегодня являются одним из важнейших средств спортивной педагогики[[1]](#footnote-1)[[2]](#footnote-2)[[3]](#footnote-3). В условиях значительно увеличившегося потока данных очень важно переходить к новым средствам их обобщения для более эффективной их обработки[[4]](#footnote-4). При управлении процессом спортивной подготовки и анализе соревновательной деятельности зачастую присутствует высокий уровень субъективности и очень остро стоит задача повышения объективности оснований принятия решений для улучшения их качества, формирования эффективных прогнозов и управления на основе прогнозирования[[5]](#footnote-5)[[6]](#footnote-6). Поэтому востребованными могут быть математические модели в которых отражается объективная связь между целевым, как правило соревновательным упражнением, которое выступает в роли зависимой переменной и одним или несколькими упражнениями, выполняемыми с целью воздействия на целевое, которые будут выступать в роли независимых переменных.

Поэтому данное пособие может быть использовано во всех практических и исследовательских задачах, в которых бакалаврам необходимо определить взаимосвязь между упражнениями в ходе спортивной подготовки или в случаях необходимости осуществить прогноз эффективности выбранного подхода в тренировочной или соревновательной деятельности.

# Применение интерполяции, аппроксимации и экстраполяции в изучении спортивной деятельности

Традиционно в выпускных квалификационных работах, посвящённых теории и методике единоборств используемый математический аппарат, относится к статистическим методам определения различий в распределении двух величин, характеризующих изучаемый аспект спортивной деятельности. При этом за рамками изучения остаётся характер связи между величинами, возможный прогноз основанный на этой связи и более объективное описание имеющихся тенденций. Именно эти проблемные направления и могут быть закрыты с помощью регрессионного анализа.

Существует несколько сходных между собой математических понятий: интерполяция, аппроксимация и регрессия. У них общая цель: из семейства функций выбрать ту, которая обладает определенным свойством. Рассмотрим учебный набор данных состоящий из двух упражнений (рис.1)

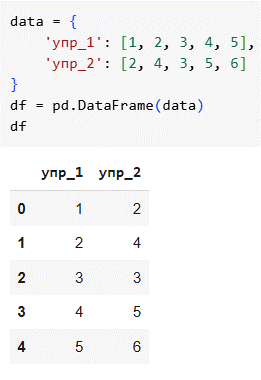


Рис.1 – Создание учебного набора данных

Полученный набор данных достаточно прост для понимания. Упражнение №1 – это вспомогательное упражнение, которое представляет из себя простой числовой постоянно растущий ряд. Это наша независимая переменная. Предполагаем, что спортсмен совершенствуясь может выполнять его в большем количестве на каждом занятии. Упражнение №2 – это целевое упражнение, изменение которого желательно понять для последующего планирования учебно-тренировочного процесса и управления им. Это наша зависимая переменная, так как нами рассматривается изменение второго упражнения в зависимости от изменения первого. Первое, что необходимо понять – это каким образом, по какому закону. Изменяется наше целевое упражнение. Это достаточно сложная задача, так как мы имеем замеры только в нескольких точках, а то, что происходит между этими известными нам значениями остаётся неизвестным. Операция, позволяющая представить набор чисел в виде функциональной зависимости называется интерполяцией.

Другими словами, интерполяция представляет из себя способ выбрать функцию, которая проходит через определённые нами заранее точки. Для интерполяции используют полиномы Лагранжа, сплайны, интерполяция может быть многомерной билинейной или трилинейной. Может быть использован метод ближайшего соседа и другие. При этом несмотря на то, что избираемая функция проходит через все известные точки остаётся вопрос о качестве восстановления зависимости в неизвестных промежутках и решение об её характере остаётся достаточно субъективным. Приведём самый простой пример, мы построим ступенчатый график кусочно-постоянной функции, считая, что измеренные значения остаются неизменными некоторое время, а затем изменяются максимально быстро – «ступенькой». Даже в таком случае мы получаем три возможных варианта (рис.2)

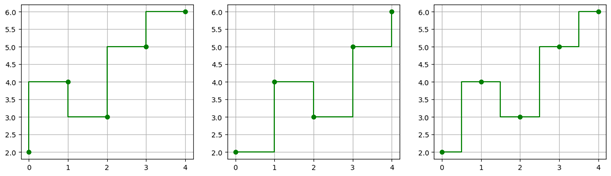


Рис.2 – Три возможности интерполяции ступенчатой функцией учебного набора данных

При кусочно-постоянной интерполяции интерполяционный многочлен на каждом отрезке  равен некоему постоянному значению (константе). В нашем случае эта константа соответствует левому, правому или арифметическому среднему измеренному нами значению.

Для правой кусочно-постоянной интерполяции значение функции , при значении аргумента , получаем:



Для левой кусочно-постоянной интерполяции значение функции, при значении аргумента , получаем:



Третьим способом будет выбор значений функции  в середине интервала между соседними измерениями и это описывается следующим образом:



Таким образом мы три различных варианта только одного из способов интерполяции имеющихся данных и выбор одного из них может обусловлен только опорой на знания предметной области, которые позволят уменьшить недостатки модели. Иным способом интерполяции может быть кусочно-линейная интерполяция. Эта возможность возникает, когда мы предполагаем, что значения нашей функции изменяются не скачком, а постепенно.

На всех интервалах аргумента интерполирующая функция представляется в виде одной из линейных функций . Значения углового коэффициента, который определяется тангенсом угла между прямой и осью абсцисс и значением свободного члена  находятся из соблюдения условий интерполяции на концах отрезка : , . С помощью этих условий можно получить систему уравнений:

,откуда находятся , . При этом острый угол (положительное значение ) обозначает линейный рост значений функции, тупой (отрицательное) обозначает уменьшение и равный нулю – отсутствие изменений.  в свою очередь является показателем ординаты точки пересечения функции с осью ординат.

Учитывая вышеизложенное, функцию  можно записать в более расширенном виде:

, если , получаем:



Полученная функция конечно является непрерывной, но не гладкой и ее производная будет не определена в каждом узле интерполяции, что может ухудшить дальнейшее использование полученной функции. Погрешность такой интерполяции конечно будет меньше, чем в случае кусочно-постоянной интерполяции, но даже здравый смысл подсказывает, что зависимость описывающая процесс спортивной подготовки должна быть гладкой и как повышение уровня подготовленности спортсмена, так и снижение этого уровня происходит не одномоментно, а растянуто во времени.

Для достижения этой гладкости и применяют для интерполяции кубический сплайн. Само слово сплайн, от английского *spline,* обозначает определённый вид гибкой линейки, используемых в черчении для создания гладких кривых через заранее заданные точки на плоскости (рис.3).



Рис.3 – Линейка сплайн (spline)

Форма подобного универсального лекала на каждом отрезке описывается кубической параболой. Сплайны используются для инженерных приложений, в компьютерной графике. Соединить три точки гладкой кривой вполне возможно с помощью полинома второй степени (параболы заданной квадратичной функцией), но если так сделать, то в точках где одна кривая переходит в другую функция будет терять гладкость и естественным решением будет использовать полином уже третьей степени, то есть в нашем случае кубический сплайн. Сплайны обеспечивают возможность задать кривые в виде нескольких наборов коэффициентов при этом точность достаточно высока.

Каждом определённом отрезке *,* *,* ищется решение в виде полинома третьей степени:



Неизвестные коэффициенты , , находятся с учётом:

* условий интерполяции: , ,;
* непрерывности функции *,*;
* непрерывности первой и второй производной:

, , .

Учитывая, что

,

для определения  неизвестных необходимо получить систему из уравнений:

, ;

, ;

, ;

, ,

где .

Недостающие два уравнения выводятся из дополнительных условий: . При этом следует заметить, что. Из системы можно исключить неизвестные , , при этом получится система из  линейных уравнений (СЛАУ) для определения коэффициентов :

, ,

, 

Иллюстрация интерполяции кубическими сплайнами и кусочно-линейной интерполяции упражнения №2 приведена на рис.4

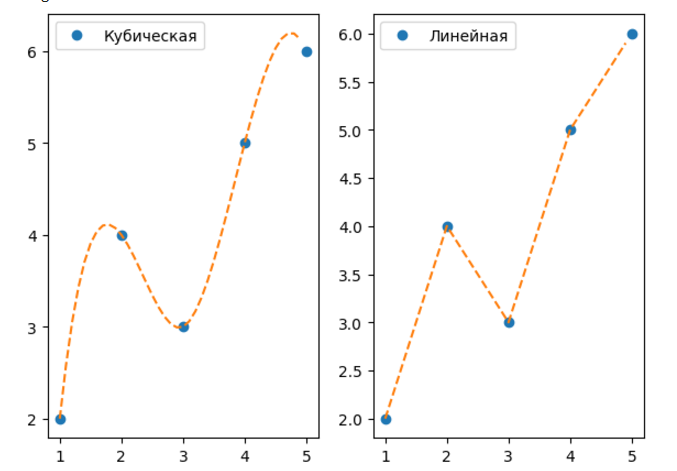


Рис.4 - Интерполяция кубическими сплайнами и кусочно-линейной интерполяция упражнения №2

При интерполировании функций используется условие равенства значений интерполяционного многочлена и выбранной функции в узлах интерполяции. Когда же исходные данные получены в результате опытных измерений, то безусловно в них присутствует некоторая погрешность. В таком случае точного выполнения условий интерполяции конечно не требуется. В этих случаях интерполирующая функция *F*(*x*) должна удовлетворять условиям , *i*=1, 2, …, N. Это условие означает, что интерполирующая функция проходит не точно через заданные точки, а в некоторой их окрестности, размер которой и определяет размер погрешности.

В этом случае речь идёт о подборе эмпирических формул,описывающих заданные точки. Построение эмпирической формулы состоит из двух этапов: подбора вида этой формулы , содержащей неизвестные параметры , и определение наилучших в выбранном заранее смысле этих параметров. Вид закономерности иногда известен из результатов проведённых ранее исследований сходных объектов, или выбирается из геометрических соображений близости. Для этого экспериментальные точки наносятся на график и путем сравнения поведения точек с графиками уже известных функций и исходя из соображений схожести подбирается общий вид зависимости. Эта схожесть в значительной степени определяется субъективно с опорой на опыт и интуицию исследователя.

Следует заметить, что интерполяция в той или иной степени точности может восстановить закономерность изменения измеряемого показателя (когда определённая функция пройдёт через все измеренные точки), но так как подобная восстановленная закономерность принадлежит только конкретному объекту в строго определённых условиях, то зачастую это не может позволить вывести наиболее общую закономерность для целого класса схожих объектов в несколько изменённых условиях. Иными словами, попытка восстановить влияние одного упражнения на другое будет верной только для одного спортсмена, только в строго определённых условиях и переносить выявленную закономерность на другого спортсмена, когда обстоятельства тренировочного процесса также претерпели изменения будет малопродуктивно.

Для получения именно обобщённого представления, выявления тенденций в изучаемом явлении необходима уже аппроксимация. Она представляет собой способ выбрать из «более простых» функций лучшее приближение для «более сложной» функции на определённом заранее выбранном промежутке, с выбором точно определенного предела возможной ошибки. Аппроксимацию используют, если существует потребность получить функцию, максимально похожую на имеющуюся, но более удобную для вычислений и последующего анализа (дифференцирования, интегрирования и т.д.). Безусловно закономерность, полученная при аппроксимации более устойчива и продуктивна при обобщении. Например, когда необходимо данные полученные на выборке распространить на всю генеральную совокупность, небольшой группы спортсменов на остальных в данном виде спорта.

Для практики наиболее частым является случай аппроксимации функции многочленами, т.е. .

Используя имеющийся набор данных можно попытаться подобрать полиномы второй и третьей степени, экспоненциальную и тригонометрическую функции. Всего пять различных функций, для наглядности используем коэффициенты обозначенные a, b, c:



Визуально результаты подбора функции к исходному набору данных можно видеть на рис.

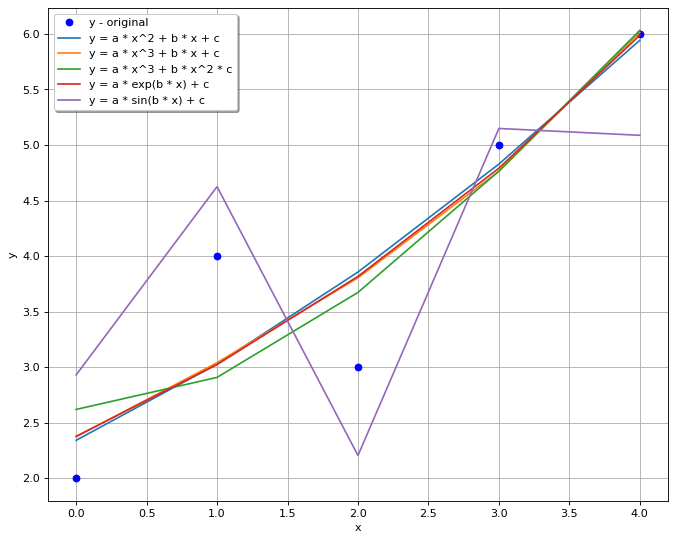


Рис.5 – Результаты аппроксимации упражнения №2 пятью различными функциями.

Визуально несколько сложно определить какая из подобранных функций наиболее точно соответствует имеющемуся набору данных и для численного выражения этого соответствия можно использовать коэффициент аппроксимации (детерминации) R2. Данный коэффициент вычисляется по формуле:



Он показывает насколько хорошо выбранная функция «объясняет» имеющийся набор данных, вернее из единицы вычитается отношение дисперсии случайной ошибки к общей дисперсии «объясняемых» данных. Говоря иначе это доля объяснённой суммы квадратов в общей сумме. Для двух переменных этот коэффициент равен квадрату коэффициента корреляции. Значения продуктивности выбранных функций для аппроксимации имеющихся данных представлены в таблице 2.

Таблица.2 – Оценки продуктивности аппроксимации результатов упражнения №2 различными функциями

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Функция № | R2 | Значения зависимой переменной при измерении (для исходных данных) или вычислении (для аппроксимации функцией) | | | | |
| Исх. | 1 | 2 | 4 | 3 | 5 | 6 |
| 1 | 0.82 | 2.34285714 | 3.02857143 | 3.85714286 | 4.82857143 | 5.94285714 |
| 2 | 0.82 | 2.375 | 3.04166667 | 3.80555556 | 4.76388889 | 6.01388889 |
| 3 | 0.79 | 2.62052877 | 2.9088647 | 3.67371695 | 4.76485225 | 6.03203732 |
| 4 | 0.82 | 2.37816399 | 3.02445753 | 3.81786864 | 4.79188607 | 5.98762174 |
| 5 | 0.73 | 2.93140483 | 4.62573871 | 2.20601159 | 5.14925541 | 5.08768085 |

Как можно заметить R2 изменяется от 0 до 1 и чем ближе к единице, тем лучше функция «объясняет» имеющиеся данные, при этом значения коэффициента большее 0,5 можно считать приемлемым для подбора. В имеющемся случае можно заметить, что оценки, превышающие значение 0,8 достаточно высоки и большинство функций, справились с задачей и теперь уже исследователь может самолично выбрать любую из них в зависимости от собственных предпочтений и планов по дальнейшему использованию полученной закономерности.

Задача объяснения наблюдаемого явления может считаться выполненной средствами интерполяции и аппроксимации. Остаётся нерешённой задача использования полученных моделей для работы с новыми, ранее неизвестными данными и здесь необходима экстраполяция. Она представляет собой метод нахождения значений, находящихся за пределами изученного интервала. Как и в случае применения интерполяции, существует достаточно большое количество методов для расчета экстраполяции. Многие из них совпадают с методами интерполяции.

Существует несколько основных подходов для применения экстраполяции в спортивной практике.

Экстраполяция может осуществляться на основе индукции (это так называемая тенденциальная экстраполяция) основана на вполне естественном предположении, что имеющиеся тенденции будут сохраняться с течением времени. Например, если увеличение объёма или интенсивности одного из упражнений влияет на другое, то можно предположить, что эта тенденция сохранится и дальше.

Экстраполяция с применением аналогового моделирования заключается в апеллировании к сходным ситуациям в прошлом для прогнозирования будущих состояний объектов. Например, есть данные о том, какие результаты были достигнуты определённым контингентом при использовании конкретных средств и методов, то можно при совпадении исходных условий прогнозировать сходное развитие событий.

Экстраполяция выборочных данных на всю генеральную совокупность наиболее часто применяемая в спортивной науке. Метод предполагает, что можно сделать выводы о всей совокупности, изучив только небольшую ее часть. Традиционно используется экспериментальная методика для подготовки группы спортсменов и по результатам оценки применения делается вывод о возможности расширения применения методики для всех спортсменов с подобными характеристиками.

Таким образом интерполяция, аппроксимация и экстраполяция имея решая различные задачи часто используют схожие методы и все они в некоторой мере используются при построении регрессионных моделей.

**Вопросы для контроля.**

Сущность, методы и основание для проведения интерполяции

Особенности кусочно-линейной интерполяции

Сущность интерполяции кубическими сплайнами

Сущность, методы и основание для проведения аппроксимации

Оценка качества аппроксимации с помощью коэффициента R2

Сущность и методы проведения экстраполяции

# Теоретико-методические основы регрессионного анализа спортивной деятельности

Регрессия в свете вышеописанного может быть представлена как «интерполирующая аппроксимация с возможностью экстраполяции». Линия (точнее гиперплоскость, иными словами поверхность размерность, которая на единицу меньше моделируемой) функции должна проходить не обязательно через, но обязательно как можно ближе к точкам, значения которых получены в ходе эксперимента и при этом оставаться максимально простой чтобы можно было определить наиболее общую тенденцию, которая в последующем будет использована для прогнозирования. Более точно регрессию можно определить, как способ выбрать из определённого семейства функций такую, которая будет минимизировать заранее определённую для этого функцию потерь. Функция потерь в свою очередь отражает то насколько сильно подбираемая функция отклоняется от значений заранее измеренных точек исследуемого набора данных. Следует не забывать, что так как точки исследуемого набора данных получены в результате эксперимента, то эти экспериментальные данные неизбежно содержат различные неточности, ошибки измерений, дополнительный шум, поэтому и требуется, чтобы функция передавала именно общую тенденцию, а не точно проходила через все измеренные точки.

Наиболее простым видом регрессионной модели является линейная. Линейные регрессионные модели как видно из названия предполагают, что определяемый критерий линейно зависит от признаков, описывающих объект или процесс. Иными словами, модель представлена в виде суммы переменных, умноженных на определённый коэффициент:



Первая в мире публикация, посвящённая методу регрессии, принадлежит Адриен Мари Лежандру (1805 год), хотя считается, что Иоганн Карл Фридрих Гаусс пришел к данному методу раньше так как успешно использовал регрессию для предсказания орбиты Цереры карликовой планеты, считая её кометой.

Достоинства линейных регрессионных моделей:

1. Меньшие требования к вычислительным мощностям, а значит скорость и простота получения модели.
2. Хорошая интерпретируемость модели. Линейная модель является интуитивно понятной и применимой в практической деятельности. По полученным коэффициентам регрессионной модели можно точно сказать, как каждый оцениваемый фактор влияет на итоговый результат, зависимую переменную. Это позволяет делать на основе значений вычисленных коэффициентов полезные выводы о планировании спортивной деятельности.
3. Широкая применимость. Большое количество реальных процессов в спорте, экономике и биологии можно с достаточной точностью описать различными линейными моделями.
4. Хорошая изученность линейных регрессионных моделей. Для линейной регрессии давно известны обстоятельства, ухудшающие продуктивность моделей (например, мультиколлинеарность) и соответственно выработаны способы решения возникающих проблем, разработаны и реализованы тесты оценки.

Недостатки линейных регрессионных моделей:

1. Линейные регрессионные модели не предназначены для выявления сложных зависимостей в данных.
2. При наличии мультиколлинеарности, то есть линейной взаимосвязи между объясняющими, независимыми переменными, что приводит к неустойчивости оценок параметров модели, что в свою очередь выражается в увеличении статистической неопределенности. Возникает дисперсия оценок и результаты оценки могут сильно различаться для различных выборок несмотря на то, что они являются однородными

Как говорилось ранее, применяемый в математическом моделировании регрессионный анализ призван дать определённые оценки взаимосвязей между зависимой переменной (в нашем случае это результат целевого/соревновательного упражнения, значение параметра соревновательной деятельности или просто «результат») и одной или несколькими независимыми переменными (для нас вспомогательными упражнениями, параметрами соревновательной деятельности иногда называемыми «характеристиками», «предикторами» или «объясняющими переменными»). Наиболее простым и применимым на практике в регрессионном анализе является линейная регрессия, при которой определяется прямая (в случае если исследуется временной ряд, то линия тренда), которая наиболее точно описывает полученные данные в соответствии с определенной заранее функцией потерь (рис.6).

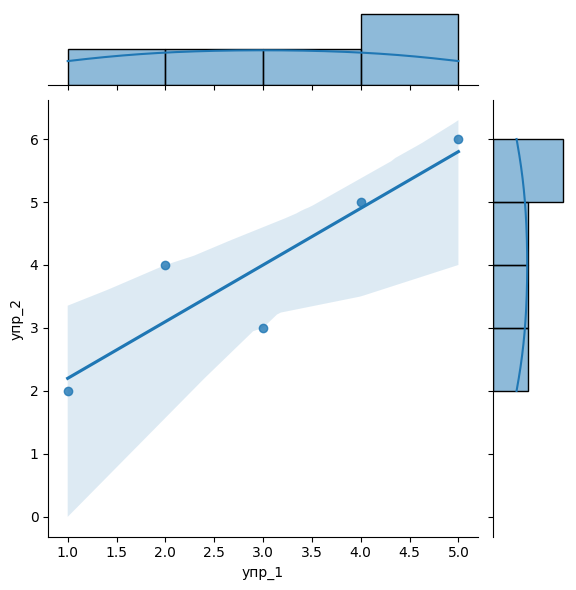


Рис. 6 – Пример линейной регрессии на учебном наборе данных (линия отражает линейную зависимость количества выполнений упражнения №2 от упражнения №1)

Функция потерь призвана как раз определить величину различия между имеющимся набором данных и прямой описываемой полученным линейным уравнением. Одним из самых широко употребимых для построения регрессионных моделей является метод обычных наименьших квадратов который вычисляет уникальную гиперплоскость, основываясь на определённом минимуме суммы квадратов отклонений, вычисленных и экспериментальных данных определяя их степень близости друг к другу**.**

Рассмотрим построение регрессионной модели методом наименьших квадратов (МНК).

Можно сказать, что для эмпирически полученных данных  выбирается вид функциональной зависимости:  где коэффициенты являются для исследователя неизвестными и должны быть определены. В данном случае для нахождения неизвестных коэффициентов в качестве функции потерь используется метод наименьших квадратов. Сумма квадратов отклонений между вычисленными по эмпирической формуле и заданными опытными данными определяется по следующей формуле:

.

Здесь описывается разность между эмпирическими и вычисленными значениями интересующих нас переменных возведённая в квадрат. Возведение в квадрат необходимо для избавления от знака, возникающего при разности. Эта разность может быть, как положительной, так и отрицательной и при сложении положительные и отрицательные погрешности уравновесят друг друга, искажая представление об ошибке. Так как при возведении в квадрат действительного числа получается только положительное значение вне зависимости от того было ли число, возводимое в степень положительным или отрицательным, то можно получить объективную информацию об ошибке, погрешности модели не только в каждой точке, но и на всем наборе данных в целом. Параметры  находятся при достижении минимума функции , поэтому используется именно подход с возведением в квадрат, так как выбор например модуля разности в виде функции потерь не позволит легко её дифференцировать и соответственно будет затруднительно найти минимум функции.. В нахождении минимума функции суммы квадратов отклонений и состоит метод наименьших квадратов (МНК).

Известно, что для нахождения экстремума в котором и находится минимум функции все частные производные от  по параметрам  равны нулю:



Традиционно в качестве эмпирической функции рассматривается полином некоторой степени:

.

В этом случае формула для определения суммы квадратов отклонений будет иметь следующий вид:



При вычислении частных производных по каждому из параметров будут получаться следующие формулы:







Если приравнять полученные выражения нулю так как нужно получить точку экстремума в данном случае соответствующий минимуму функции, и собрать коэффициенты при неизвестных , то можно получить следующую систему линейных уравнений:



.



Данная система уравнений имеет название нормальной системы**.** Если произвести решение полученной системы линейных уравнений то будут найдены неизвестные коэффициенты .

В случае исследования полинома первой степени , который выглядит следующим образом: , система нормальных уравнений примет вид:



.

В случае полинома второй степени при  количество уравнений будет увеличено на одно и далее подобная тенденция сохраняется, при увеличении степени (порядка) полинома необходимо будет решить всё более сложную систему уравнений:

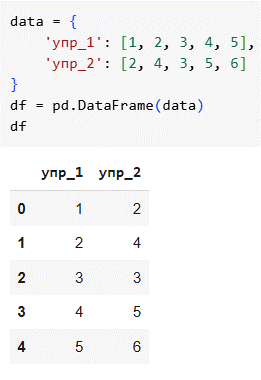


.



Как правило, выбирают несколько эмпирических зависимостей. По МНК находят коэффициенты этих зависимостей и среди них находят наилучшую по минимальной сумме отклонений.

Подставляя значения учебного набора данных

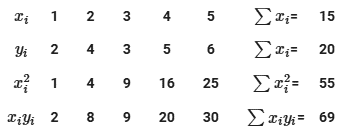


В полученную систему уравнений для построения линейной регрессионной модели:





Получим следующие результаты вычислений:



Подставим вычисленные значения в полученную систему уравнений:



Полученную систему можно решить различными способами, но для наглядности используем известный со средней школы метод и выразим из первого уравнения одно неизвестное через другое:



Потом подставим полученное выражение во второе уравнение и получим значение одной из неизвестных:



Теперь зная значение одного из параметров можно легко вычислить и второй:



Таким образом мы без использования сложных вычислений получили линейное уравнение описывающее зависимость одного упражнения от другого, иными словами линейную регрессионную модель спортивной тренировки, состоящей всего из двух упражнений:



Визуально данную модель можно оценить на рис.7

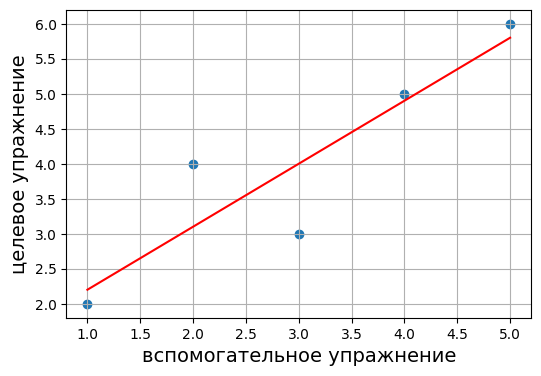


Рис.7 – Визуализация линейной регрессионной модели, построенной на учебных данных

Следует помнить, что данная модель действует только на определённом интервале и распространять её на произвольный интервал может быть опрометчиво. После определения характера зависимости и коэффициентов модели необходимо проверить её продуктивность и ранее мы использовали коэффициент аппроксимации (детерминации) R2. В случае с регрессионной моделью этот метод так же применим и в исследуемом случае R2 = 0.81, что является достаточно высоким показателем и находится на одном уровне с проведёнными ранее аппроксимациями. Для оценки функции потерь, в качестве которой использовалась среднеквадратичная ошибка прогноза, то для этого используем метрику RMSE (Root Mean Squared Error), представляющей собой корень квадратный из вычисленной ошибки. Это сделано для того, чтобы приблизить оценку к единицам измерения переменных. В исследуемом случае RMSE = 0.61. Таким образом можно условно заявить, что продуктивность модели составила более 80%, а ошибка меньше одной единицы.

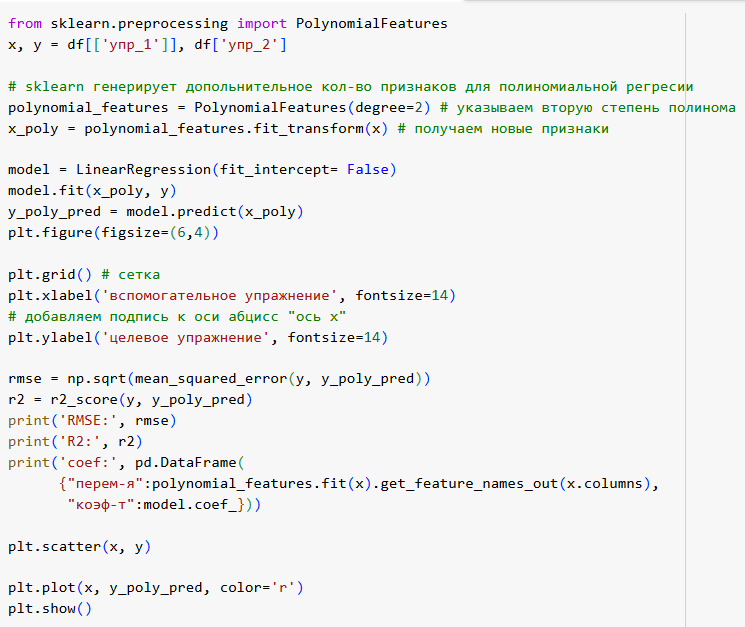


Рис.8 –Код построения полиномиальной регрессионной модели с визуализацией и оценкой продуктивности

Продолжая исследование можно построить уже модель полиномиальной регрессии и степень полинома повысим на единицу, чтобы она была равна двум, что соответствует параболе. Опуская вычисления сразу воспользуемся возможностями библиотеки алгоритмов машинного обучения Scikit-Learn языка программирования Python.

В результате применения библиотеки Scikit-Learn была сформулирована регрессионная модель, описываемая полиномом второй степени:



Визуализацию полиномиальной регрессионной модели второй степени можно видеть на рис.9

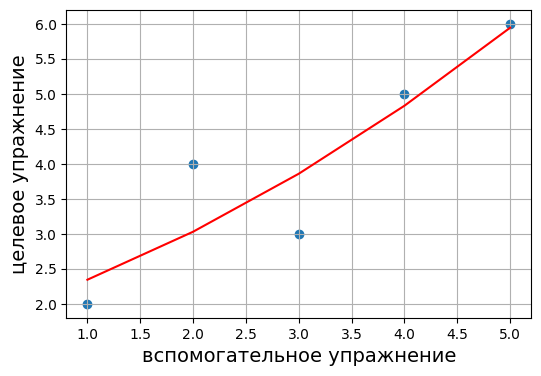


Рис.9 – Визуализация регрессионной модели, описываемая полиномом второй степени, построенной на учебных данных

Продуктивность созданной модели, определяемая коэффициентом детерминации R2 = 0.82 и функцией потерь RMSE = 0.61 незначительно лучше линейной модели, но при этом объяснительная способность полиномиальной модели гораздо хуже, чем у интуитивно понятной линейной модели, поэтому вполне обоснованно использовать более простую и понятную модель.



Визуальное отображение полиномиальной регрессионной модели второй степени можно видеть на рис.10

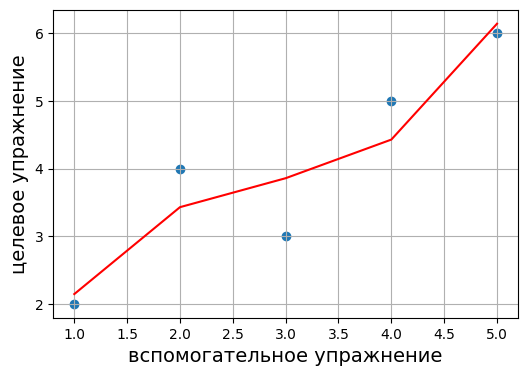


Рис.10 – Визуализация регрессионной модели, описываемая полиномом третьей степени, построенной на учебных данных

Продуктивность созданной модели, определяемая коэффициентом детерминации R2 = 0.86 и функцией потерь RMSE = 0.54. В этом случае продуктивность модели уже значительно выше, но описываемая зависимость становится излишне сложной для понимания и потому малоприменимой для спортивной практики.

Если же, например, степень полинома, которым описывается учебный набор данных увеличить до 5, то функция пройдёт через все описываемые точки (рис.11)



Рис.11 – Визуализация регрессионной модели, описываемая полиномом пятой степени, построенной на учебных данных

При этом продуктивность созданной модели, определяемая коэффициентом детерминации R2 = 1 и функцией потерь RMSE приближающейся к нулю может служить описанием сложившейся ситуации, но модель при этом теряет устойчивость и при введении в неё новых данных. Это происходит в том числе и потому, что некоторые методы, в том числе и метод наименьших квадратов чувствительны к выбросам. В машинном обучении такая ситуация называется переобучением, а модель соответственно считается переобученной. Это происходит потому, что различные погрешности, возникающие при эмпирическом исследовании, принимаются моделью за точное описание и вместо имеющихся в исследуемых процессах закономерностях описываются в том числе и эти погрешности, называемые «шумом». Даже отвлекаясь от проблемы переобучения, полином пятой сложности очень сложен для понимания процессов и поэтому мало применим на практике.

Подводя итоги можно напомнить, что регрессионный анализ очередь используется в первую для достижения двух концептуально различных целей. Регрессионный анализ может быть продолжением традиционного для спортивной науки сравнения с использованием статистических критериев и использован для определения значимых причинно-следственных связей между группой независимых переменных и зависимой переменной. Другое применение регрессионного анализа происходит при прогнозировании и выявлении общих тенденций в изучаемом наборе данных. Чтобы использовать регрессии для прогнозирования или определения причинно-следственных связей, исследователь должен тщательно обосновать, почему существующие взаимосвязи обладают предсказательной силой в новом контексте или почему взаимосвязь между двумя переменными имеет причинно-следственную интерпретацию. Последнее особенно важно, когда исследователи надеются оценить причинно-следственные связи, используя данные наблюдений.

**Вопросы для контроля.**

Сущность и основание для проведения регрессионного анализа

Основные преимущества регрессионных моделей

Основные недостатки регрессионных моделей

Построение регрессионной модели методом наименьших квадратов

# Использование регрессионного анализа для проведения исследований спортивной подготовки (на примере технико-тактической подготовки).

В качестве примера использования исследований регрессионного анализа в исследовании различных аспектов спортивной подготовки можно привести описание исследований, проводимых на кафедре «Теории и методики единоборств» РУС «ГЦОЛИФК». Мы подвергли изучению бросок «подхватом изнутри». Данный бросок пользуется достаточно высокой популярностью у спортсменов на соревнованиях различных уровней[[7]](#footnote-7)[[8]](#footnote-8)[[9]](#footnote-9). В качестве испытуемых выступили студенты кафедры специализаций, служебно-прикладные единоборства и дзюдо квалификациями кандидат в мастера спорта и мастера спорта. Использование спортсменов различных специализаций возможно в этом случае так как исследуемый бросок присутствует в различных видах единоборств.

Неудивительно, что при значительной популярности и зрелищности данного броска его исследованию уделяли внимание его изучению в целях совершенствования для достижения более высокого спортивного результата. Достаточно часто внимание исследователей направлено на биомеханические характеристики движения броска и в основном исследуются вопросы устойчивости во время выполнения броска[[10]](#footnote-10)[[11]](#footnote-11)[[12]](#footnote-12)[[13]](#footnote-13).

Менее изученным является изучение динамики нижних конечностей и туловища во время выполнения броска. Другим важным вопросом является связанная с вышеназванной проблема классификации этого приёма. Так бросок подхватом изнутри относится к броскам, выполняемым в первую очередь за счёт движения ногами. Несмотря на это существуют также мнения высококвалифицированных специалистов, что данный бросок необходимо относить к броскам, выполняемым преимущественно движением туловища[[14]](#footnote-14). Так же другие корейские и японские исследователи утверждают о возможности отнесения данного броска к броскам выполняющимися в основном туловищем на основании схожести с броском через бедро и ставят вопрос о высокой значимости наклона и особенно амплитуды наклона туловища спортсмена при выполнении броска подхватом изнутри[[15]](#footnote-15)[[16]](#footnote-16)[[17]](#footnote-17).

Поэтому для исследования биомеханических особенностей данного приёма ограничившись значениями амплитуды движения в тазобедренных суставах, которые могут служить как показателем активности, атакующей (маховой) ноги, так и наклона туловища. Бросок подхватом выполнялся так что, правая нога избиралась всеми испытуемыми в качестве атакующей (маховой), а левая соответственно опорной к которой выполнялся наклон. Исходя из этого изменения значения угла между туловищем и бедром правой ноги определяли динамику махового движения атакующей ноги, а изменение значения угла между туловищем и левой ногой определяет динамику движения туловища. Для определения параметров исследуемого движения, а именно значения амплитуды движения в тазобедренных суставах испытуемых спортсменов нами был использован разработанный нами программно-аппаратный комплекс с использованием искусственного интеллекта, компьютерного зрения[[18]](#footnote-18) (отражено на рис.12)

Рис.12 – Пример выполнения броска с дополнительной обработкой видео системы компьютерного зрения.

Получив данные и убедившись в их полноте, мы провели первичный визуальный анализ. Первым шагом было построение линейного графика зависимости амплитуды движений в левом и правом тазобедренных суставах от времени при выполнении серии из четырёх последовательных имитаций броска (рис.13). При ближайшем рассмотрении можно заметить, что высока вероятность существования взаимосвязи между исследуемыми параметрами.

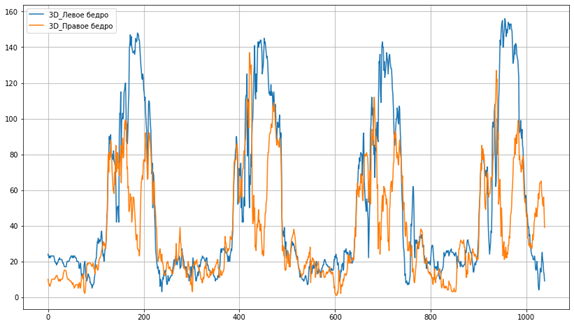


Рис.13 – Временая зависимость амплитуды движений в левом и правом тазобедренных суставах выполнения серии из четырёх последовательных имитаций броска

Более углубленное понимание возможно при расмотрении в размаха и распределение вероятностей динамики нижних конечностей при выполнении броска (рис.14)

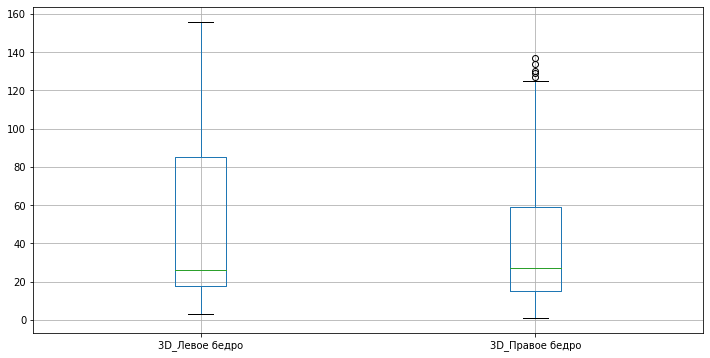


Рис.14 – Размах и распределение вероятностей динамики нижних конечностей при выполнении подхвата изнутри

Каждое выполнения приёма безусловно различно, но во всех из них при этом движение наклона туловища по амплитуде превышает амплитуду атакующего движения ногой, которое в свою очередь согласно наблюдаемым выбросам на графике является более вариативным. Имеющиеся же оценки требуют объективной оценки, описательные же статистики выполняемой серии бросков представлены в таблице 3.

Таблица.3 – Описательные статистики исследуемого набора данных

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1-й бросок | | 2-й бросок | | 3-й бросок | | 4-й бросок | |
| Левое бедро | Правое бедро | Левое бедро | Правое бедро | Левое бедро | Правое бедро | Левое бедро | Правое бедро |
| Среднее значение | 85.81 | 58.24 | 82.18 | 60.72 | 76.70 | 56.69 | 89.47 | 58.45 |
| Стандартное отклонение | 40.39 | 23.43 | 41.35 | 31.38 | 40.13 | 23.89 | 45.60 | 25.20 |
| Минимум | 20 | 15 | 18 | 15 | 7 | 8 | 21 | 21 |
| Процентиль 25% | 48.25 | 38 | 42 | 31 | 45 | 37.75 | 49 | 34 |
| Процентиль 50% | 87.5 | 63.5 | 87 | 56 | 80 | 59 | 85 | 58 |
| Процентиль 75% | 116.75 | 77 | 115 | 87 | 106.25 | 76.5 | 140 | 77 |
| Максимум | 148 | 99 | 145 | 137 | 143 | 112 | 156 | 127 |

Как можно заметить присутствуют видимые различия не только между значениями углов между туловищем левым и правым бедром в каждом из серии четырёх бросков, но и видятся некоторые различия между отдельными бросками согласно значениям оцениваемых углов. Возникает вопрос о значимости наблюдаемых различий. Для установления этой значимости нами был произведён дисперсионный анализ и в качестве критерия был использован H-критерий Краскела — Уоллиса. При сравнении углов наклона в серии приёмов H-критерий = 3.5 (p=0.2), а при оценке амплитуды атакующей (маховой) ноги H-критерий = 0.9 (p=0.7). Как можно заметить, при соблюдении традиционного подхода к значимости при р≤0,05 выявленные нами различия не могут считаться значимыми. Это значит, что несмотря на видимые различия между бросками в серии мы можем считать их неразличимыми, что говорит о стабильности техники выполнения приёма. В тоже самое время различие между амплитудами наклона и атакующей ноги по H-критерию Краскела — Уоллиса равен 71.5 (р≤0,001), что в свою очередь говорит о высоком уровне различий. Соответственно можно установить зависимость между движениями атакующей ноги и наклона туловища. Поэтому проведём корреляционный анализ.

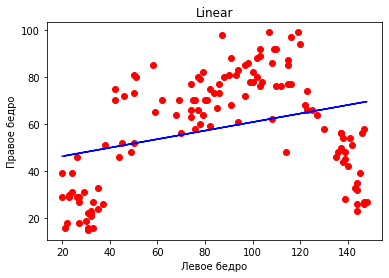


Рис.15 – Визуализация простой регрессионной (линейной) модели на исследуемом наборе данных

Коэффициент корреляции оценивающий амплитуд движений ногой и туловищем оказался достаточно низким (r = 0.32). Следуя из этого построение линейных регрессионных моделей может быть малоперспективным. Пример простой линейной модели, представленный на рисунке 15 наглядно показывает, что линейная зависимость недостаточно хорошо описывает исследуемый набор данных.

Осознав, что зависимость нелинейная, мы использовали полиномиальную регрессию описывая имеющуюся взаимосвязь полиномом второй степени (рис.16).

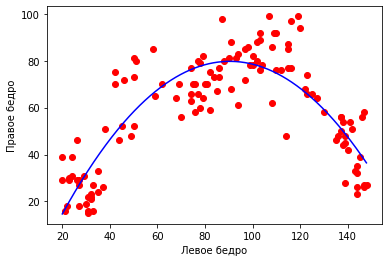


Рис.16 - Визуализация регрессионной модели исследуемого набора данных полиномом второй степени

Визуально можно заметить, что полином второй степени гораздо лучше описывает изучаемый набор данных. Остаётся вопрос о выборе модели наиболее соответствующей целям спортивной деятельности. Выбор модели традиционно происходит путём сравнения продуктивности моделей путём сравнения коэффициентов детерминации и значение функции потерь. Для понимания продуктивности полиномиальных моделей мы провели множественное моделирование последовательно изменяя степень полинома со второй до девятнадцатой.

Последовательно меняя значения степени полинома, мы видим, что наилучшую продуктивность показывает модель, выраженная уравнением шестой степени (табл.4).

Таблица.4 - Изменение продуктивности модели и среднеквадратической ошибки в зависимости от степени используемого полинома.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Степень полинома | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Продуктивность модели (R2) | 0.73 | 0.75 | 0.75 | 0.75 | **0.80** | 0.79 | 0.78 | 0.78 | 0.79 |
| Среднеквадратическая ошибка (MSE) | 12.22 | 11.72 | 11.65 | 11.63 | **10.36** | 10.81 | 11.04 | 10.86 | 10.82 |
| Степень полинома | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| Продуктивность модели (R2) | 0.76 | 0.73 | 0.70 | 0.67 | 0.58 | 0.55 | 0.52 | 0.50 | 0.47 |
| Среднеквадратическая ошибка (MSE) | 11.33 | 12.16 | 12.88 | 13.40 | 15.10 | 15.61 | 16.10 | 16.56 | 16.98 |

Как можно заметить наиболее продуктивной и показывающей наименьшую ошибку оказалась модель, описываемая уравнением шестой степени (рис.17)

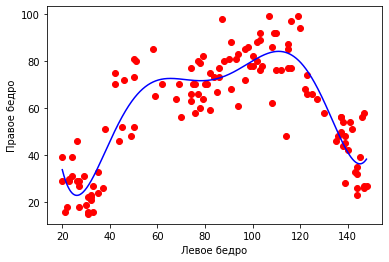


Рис.17 - Визуализация регрессионной модели исследуемого набора данных полиномом второй степени

Однако настолько сложная зависимость чрезвычайно тяжела для понимания и соответственно использования на практике. Поэтому для оценки взаимосвязи динамики наклона туловища при выполнении подхвата изнутри вполне можно производить оценку обычным квадратным уравнением:

**y** = -0.01311694**x2 +** 2.37268614**x**

где у – это динамика атакующего движения маховой ноги, а х – это динамика наклона к левой ноге.

В результате построения модели можно сделать вывод о том, что в приёме бросок подхватом изнутри на старте скорость маха ноги в два раза превышает скорость наклона туловища и в конце маха эти скорости сравниваются. Данное наблюдение может позволить, сделать вывод о том, что классификация приёма прежде всего зависит от качества, по которому производится. В случае броска подхватом изнутри мах ноги требует быстроты, в тоже время наклон туловища прежде всего силы. Кроме вопросов классификации — это наблюдение позволяет управлять подбором как объёма-интенсивности нагрузки, так и выбором упражнений в комплексе. Это позволит не просто максимизировать нагрузку, но подбирать оптимальное их значение используя индивидуально построенные модели для каждого конкретного спортсмена.

Кроме прочего если исследователь может принять решение, что выбросы не являются значимыми то может применить различные методы для усреднения, например, экспоненциальное сглаживание

Экспоненциально взвешенное скользящее среднее (EWMA) — это тип скользящего среднего (MA), который придает больший вес недавним наблюдениям и меньший вес наблюдениям, отдалённым по времени. Этот тип скользящего среднего обычно используется в экономических моделях для сглаживания краткосрочных колебаний и выделения долгосрочных тенденций или циклов. Пример обработки исследуемого набора данных с помощью экспоненциального сглаживания на рисунке 18.

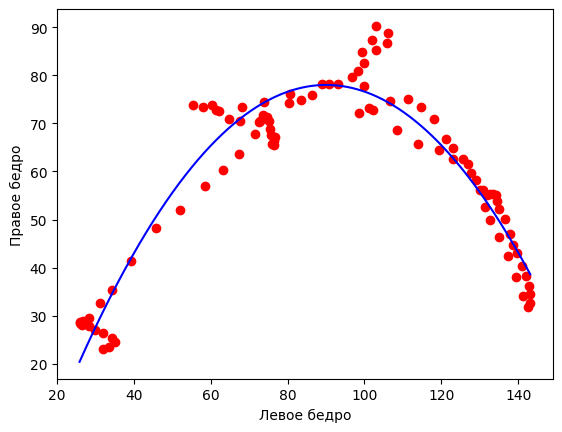


Рис.18 - Визуализация регрессионной модели исследуемого набора данных полиномом второй степени после экспоненциального сглаживания.

Получившаяся модель как можно заметить мало отличается от исходной:

**y** = -0.01397172**x2 +** 2.51651746**x**

Несмотря на столь малые различия в формуле, сравнивая продуктивности моделей можно заметить значительное её улучшение после обработки. До сглаживания коэффициент детерминации R2 = 0.73, среднеквадратическая ошибка (MSE) = 12.22. после проведения сглаживания коэффициент детерминации R2 = 0.91, среднеквадратическая ошибка (MSE) = 5.68. Такое значительное повышение продуктивности при незначительном изменении самих параметров модели делает в данном случае обработку оправданной, а модель надёжной.

**Вопросы для контроля.**

Постановка задачи для проведения регрессионного анализа ТТП

Визуальный анализ

# Сравнительный анализ соревновательной деятельности с помощью регрессионных моделей

Значительную помощь построение регрессионных моделей может помочь в анализе соревновательной деятельности. Как правило для контроля и оценки соревновательной деятельности используют большое количество факторов и установление зависимостей между ними становится одной из важных задач для совершенствования соревновательной деятельности.

Для анализа используем два набора данных большого размера описывающих соревновательную деятельность в смешанных единоборствах. первый «ММА All Pro MMA Fighters - UFC, Bellator, One» содержит 13322 наблюдений согласно 530 измеренным параметрам[[19]](#footnote-19) и второй «UFC/MMA Biggest Dataset with Differentials» содержит 5151 наблюдений согласно 22 измеренным параметрам [[20]](#footnote-20).

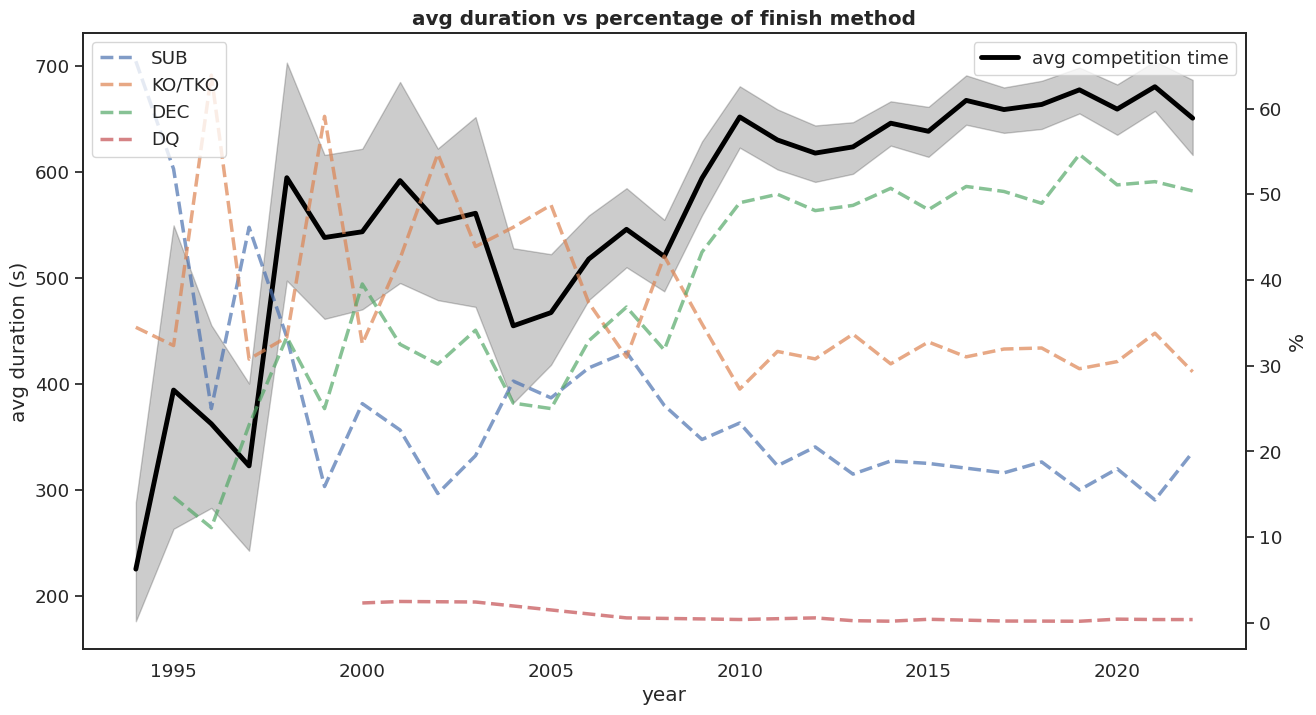


Рис.19 – Время и способ завершения боя в зависимости от года проведения соревнования

Наиболее важным кроме самой информации о победе или поражении для спортивной деятельности является способ достижения победы и соответственно причины поражения. На рисунке 19 показана зависимость способа и времени завершения боя в зависимости от года проведения состязания. Изучая график можно заметить, что время достижения победы по каждому из способов со временем стабилизировалось и можно попытаться найти важные зависимости. Наиболее важная зависимость в единоборствах – это зависимость от весовой категории спортсменов. Распределение количества боёв согласно весовой категории и способу их завершения представлено в таблице 6.

Таблица 6 – Распределение количества боёв согласно весовой категории и способу их завершения

| **Способ/**  **категория** | **DQ** | **DRAW** | **KO/TKO** | **M-DEC** | **S-DEC** | **SUB** | **U-DEC** | **Всего боёв** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Bantamweight** | 1 | 15 | 156 | 3 | 68 | 114 | 192 | 549 |
| **Featherweight** | 1 | 14 | 165 | 4 | 64 | 101 | 277 | 626 |
| **Flyweight** | 0 | 4 | 59 | 0 | 34 | 50 | 115 | 262 |
| **Heavyweight** | 3 | 12 | 341 | 9 | 28 | 96 | 144 | 633 |
| **Light Heavyweight** | 3 | 15 | 266 | 6 | 43 | 104 | 174 | 611 |
| **Lightweight** | 2 | 21 | 329 | 10 | 119 | 271 | 429 | 1181 |
| **Middleweight** | 4 | 11 | 333 | 3 | 77 | 199 | 270 | 897 |
| **Welterweight** | 5 | 17 | 377 | 10 | 112 | 216 | 425 | 1162 |

В приведённой таблице DQ обозначает дисквалификацию, DRAW обозначает победу ввиду сдачи соперника, M-DECобозначает победу ввиду решения большинством, S-DEC обозначает победу раздельным решением, U-DEC обозначает победу единогласным решением, KO/TKO обозначает победу нокаутом или техническим нокаутом и SUB обозначает победу болевым или удушающим приёмом. Ограничивая и далее выбор из соображений важности остановимся только на однозначном завершении боя в результате нокаута, болевого или удушающего приёма проведём визуальный анализ (рис.20).

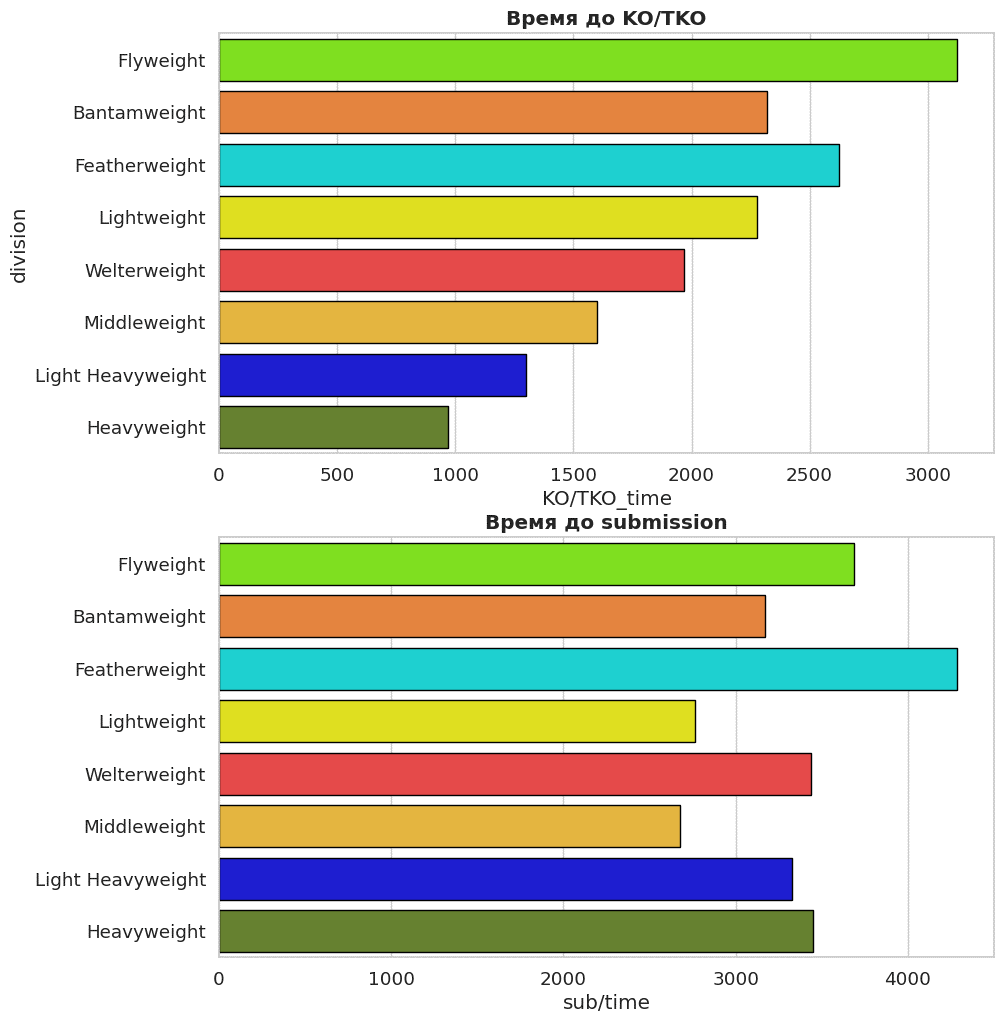


Рис.20 – Зависимость времени завершения боя нокаутом, болевым и удушающим приёмом от времени боя

Как видно из диаграмм весьма вероятность линейной зависимости времени до нокаута от веса спортсмена и менее вероятна зависимость времени до болевого или удушающего приёма. Используем библиотеку Statsmodels языка программирования Python[[21]](#footnote-21) (рис.21) Statsmodels - это универсальный пакет , который обеспечивает простые методы для получения описательной статистики, обучения моделей в первую очередь регрессионных и самое главное их оценки в удобной для оценки форме. Можно сразу получить значения большого количества тестов и критериев описывающих обученную модель без применения дополнительного кода. Statsmodels основана на экосистеме SciPy и поддерживает структуры данных в виде массивов библиотеки NumPy и структур библиотеки Pandas языка программирования Python.

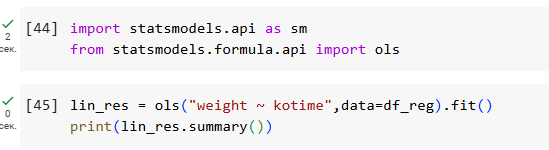


Рис.21 – Код для построения линейной регрессионной модели с использованием библиотеки Statsmodels языка программирования Python

Используя данный код построим линейную регрессионную модель зависимости времени до нокаута (технического нокаута) от веса спортсмена (рис.22). В Statsmodels реализован очень лаконичный и в тоже время информативный вывод результатов анализа данных. Линейная регрессия при выводе сообщает большое количество дополнительных оценок.

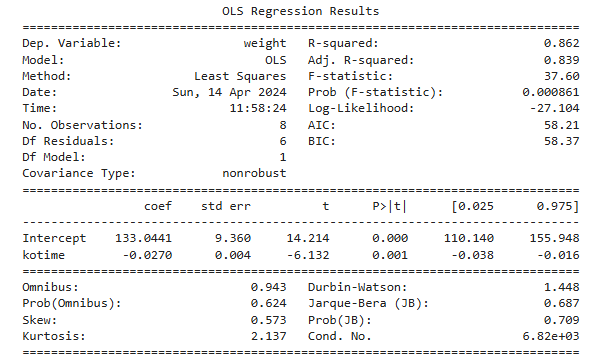


Рис.22 – Линейная регрессионная модель взаимозависимости времени до нокаута (технического нокаута) и веса спортсмена

В данном выводе можно видеть следующие обозначения:

**Dep. Variable** - имя зависимой переменной, в нашем случае это вес спортсменов.

**Model** - вид обученной модели, в нашем случае это простая линейная модель.

**Method** - используемый метод обучения, в нашем случае используется описанный выше метод наименьших квадратов.

**Date, Time** - дата обучения модели

**No. Observations**- количество наблюдений в выборке (n) , в нашем случае это восемь весовых категорий.

**Df Residuals** - степень свободы (количество наблюдений (n) - количество переменных (k) - 1).

**Df Model** - количество переменных (регрессоров) (k), в нашем случае есть только одна независимая переменная.

**Covariance Type** - по умолчанию используется ковариация, которая неустойчива к гетероскедастичности ошибки. Чтобы воспользоваться устойчивой версией, необходимо использовать гиперпараметр cov\_type=HC0/HC1/HC2/HC3. Он используется для построения матрицы ковариации параметров модели (чтобы её посмотреть, можно вызвать функцию. cov\_params()). На диагонали матриц будет стоять дисперсия оценок коэффициентов, а числа вне главной диагонали равна их ковариациям. С помощью этой матрицы мы можем сделать вывод о линейной зависимости коэффициентов регрессии друг от друга.

**Intercept** – это свободный член выводимого уравнения, когда все независимые переменные приравниваются к нулю и подбирается константа, которая наилучшим образом описывающая данные.

Вместе с Intercept описывают коэффициенты при независимых переменных (предикторах). Полученная модель может быть описана следующей формулой:

**Вес спортсмена** = -0.027**время до нокаута** + 133.0441

**std err** - дисперсия коэффициента согласно данным. Вычисляется дисперсия с помощью матрицы ковариаций согласно параметрам модели:

C=cov(β)=σ2(XXT)−1

где σ2 — это среднеквадратичная ошибка (MSE) остатков. Дисперсия каждого из коэффициентов будет стоять на диагонали данной матрицы.

**t** – это статистика t-Стьюдента, чем она больше — тем лучше измерен коэффициент (чем меньше стандартное отклонение, тем лучше коэффициент описывает зависимость). t-критерий позволяет оценить важность предикторов, предполагая, что остатки модели нормально распределены около нуля. Если остатки не ведут себя таким образом, то это говорит о наличии некоторой нелинейности между переменными и о том, что их t-тесты не следует использовать для оценки важности отдельных предикторов.

**P>|t|** - показатель значимости, например если выбран уровень < 0.05, то гипотеза о значимости коэффициента принимается при более низких значениях.

**Доверительные интервалы коэффициента** - чем меньше доверительный интервал, тем больше уверенность в значении оцениваемого коэффициента

**R-squared** - уже известный коэффициент детерминации R² (статистическая мера, представляющая долю дисперсии зависимой переменной, которая объясняется независимой переменной или переменными в модели регрессии. R2 показывает, в какой степени дисперсия одной переменной объясняет дисперсию второй переменной. Таким образом, если R2 модели составляет 0,50, то примерно половину наблюдаемых изменений можно объяснить входными данными модели) в нашем случае R² = 0.862 достаточно высок, что значит вес спортсмена и времени до нокаута имеют статистическую зависимость.

**Adj. R-squared** - скорректированный R2 используют, чтобы сравнить качество регрессионных моделей, содержащих разное количество независимых переменных. При сравнении модели с пятью независимыми переменными и модели с одной переменной, модель с пятью имеет более высокий R2. Подобная ситуация называется переобучением и может возвращать неоправданно высокое значение R2. Скорректированный R2 используется для определения того, насколько надежна корреляция и насколько она определяется добавлением независимых переменных. Скорректированный R2 настраивается на количество признаков в модели. Штраф увеличивается при добавлении каждой новой переменной. В нашем случае он равен 0.839, что не сильно отличается от значения R2 так как в нашей модели всего одна независимая переменная (предиктор).

R2, в отличие от скорректированного R2, используется для указания того, насколько хорошо регрессионная модель предсказывает ответы для новых наблюдений. Одно из ошибочных представлений о регрессионном анализе состоит в том, что низкое значение R2 — это всегда плохо.

**F-statistic** критерий Фишера можно использовать для определения общей значимости нескольких независимых переменных. Он используется для сравнения способности моделей объяснять дисперсию зависимой переменной. Если нет взаимосвязи между откликом и предикторами, F-статистика принимает значение, близкое к 1, иначе число больше. Значение F представляет собой отношение суммы квадратов средней регрессии к сумме квадратов средней ошибки. Его значение будет варьироваться от нуля до сколь угодно большого числа. В нашем случае значение равно 37, что больше единицы.

**Prob (F-statistic)** — это вероятность того, что нулевая гипотеза для полной модели верна (т. е. что все коэффициенты регрессии равны нулю). Например, если Prob(F) имеет значение 0,01, то есть 1 шанс из 100, что все параметры регрессии равны нулю. Значение будет означать, что по крайней мере некоторые из параметров регрессии отличны от нуля и что уравнение регрессии имеет некоторую достоверность (т. е. независимые переменные не являются чисто случайными по отношению к зависимой переменной).

**Durbin-Watson Statistic** «Тест Дарбина-Ватсона на автокорреляцию»– показывает вероятность того, что значения отклонения (ошибки) для регрессии имеют компонент авторегрессии первого порядка. В моделях регрессии предполагается, что отклонения ошибок некоррелированы. Небольшие значения статистики Дарбина-Ватсона указывают на наличие автокорреляции. В нашем случае автокорреляция не имеет места так как в нём только одна независимая переменная.

**Omnibus** описывает нормальность распределения наших остатков (остатки необъяснённой дисперсии), используя в качестве измерений перекос (skew) (коэффициент симметрии распределения) и эксцесс (kurtosis) (мера остроты случайной величины в горбу). 0 означает полную нормальность.

**Prob (Omnibus)** - это статистический тест, измеряющий вероятность нормального распределения остатков. Значение 1 означает совершенно нормальное распределение.

**Skew** (Перекос) - это мера симметрии распределения остатков, где 0 означает идеальную симметрию.

**Kurtosis** (Эксцесс) - измеряет остроту распределения остатков или его концентрацию около 0 на нормальной кривой. Более высокий эксцесс означает меньшее количество выбросов.

**Jarque-Bera (JB)** проверка нормальности распределения остатков по критерию Харке-Бера и его расчетный уровень значимости **Prob(JB)**. Критерий Харке-Бера является асимптотическим, расчетное значение имеет распределение хи-квадрат, поэтому данный критерий рекомендуют применять только для больших выборок

**Cond. No** Число обусловленности используется для проверки мультиколлинеарности (считается, что мультиколлинеарность вероятна, если значение Cond. No > 30). В нашем случае мультиколлинеарность не имеет места так как в нём только одна независимая переменная.

Построим модель зависимости времени до завершения боя болевым или удушающим приёмом и весом спортсменов (рис.23).

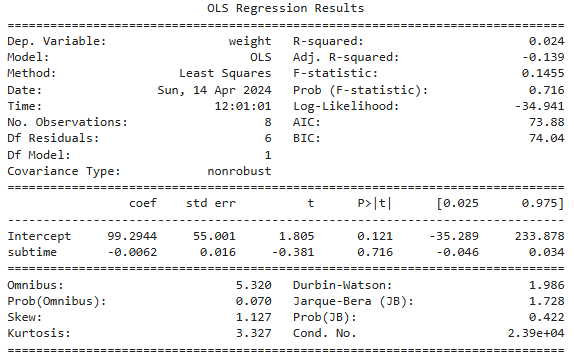


Рис.23 – Линейная регрессионная модель взаимозависимости времени до завершения боя болевым или удушающим приёмом и веса спортсмена

**Log-Likelihood** или логарифмическое правдоподобие регрессионной модели — это другой способ измерить продуктивность модели. Чем выше значение логарифмической вероятности, тем лучше модель соответствует набору данных. Значение логарифмического правдоподобия может варьироваться от отрицательной бесконечности до положительной бесконечности. Фактическое значение логарифмического правдоподобия для отдельной модели в большинстве случаев не имеет большого смысла, но оно полезно для сравнения двух или более моделей. Можно заметить, что сравнивая полученные модели в первой из них логарифмическое правдоподобие больше чем во второй (-27.104 > -34.941) соответственно первая модель лучше описывает изучаемый набор данных.

Качество статистических методов из теории информации может быть измерено информационными критериями (IC). Таким образом, это относится к методам выбора модели, основанным на функциях правдоподобия. Самый низкий балл получает лучшая модель.

**AIC** (информационный критерий Акаике) предназначен для поиска модели, которая объясняет наибольшую изменчивость данных, и штрафует модели, использующие чрезмерное количество параметров. В нашем случае первая модель имеет меньший, а значит лучший показатель AIC (58.21 < 73.88).

**BIC** (байесовский информационный критерий) из байесовской вероятности чувствителен к размеру выборки, добавляет больший штраф, более прост. В нашем исследовании необходимо предпочесть первую модель.

Для дополнительного сравнения построим модель взаимозависимости времени продолжительности боя и веса спортсмена (рис.24)

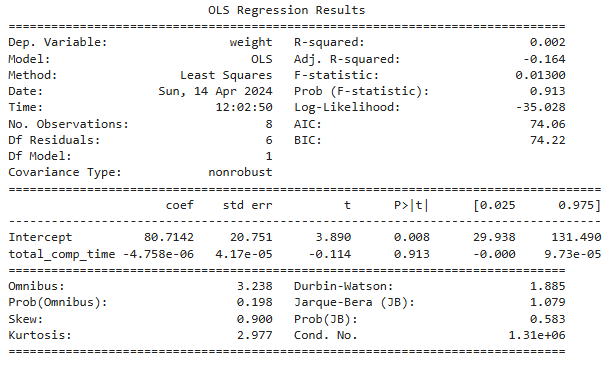


Рис.24 – Линейная регрессионная модель взаимозависимости времени продолжительности боя и веса спортсмена

Таким образом мы получили для сравнения три модели:

1. Линейная регрессионная модель взаимозависимости времени до нокаута (технического нокаута) и веса спортсмена
2. Линейная регрессионная модель взаимозависимости времени до завершения боя болевым или удушающим приёмом и веса спортсмена
3. Линейная регрессионная модель взаимозависимости времени продолжительности боя и веса спортсмена.

Результаты сравнения для большей наглядности можно объединить в таблицу 7.

Таблица 7 – Результаты сравнения полученных регрессионных моделей

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Модель | R2 | Adj.R2 | F-statistic | Prob(F-statistic) | Log-Likelihood | AIC | BIC |
| Первая | 0.862 | 0.839 | 37.60 | 0.000861 | -27.104 | 58.21 | 58.37 |
| Вторая | 0.024 | -0.139 | 0.1455 | 0.716 | -34.941 | 73.88 | 74.04 |
| Третья | 0.002 | -0.164 | 0.0130 | 0.913 | -35.028 | 74.06 | 74.22 |

В результате такого сравнительного анализа согласно значениям различных критериев можно однозначно заявить, что первая модель имеет лучшие значения коэффициентов продуктивности. Это означает, что действительно у спортсменов большего веса значительно больше вероятность завершить бой досрочно ввиду нокаута или технического нокаута. При этом при завершении боя болевым или удушающим приёмом во второй модели отсутствует статистическая взаимосвязь с весом спортсменов. Это может означать, что при проведении атак болевым или удушающим приёмом спортсмены более тяжёлых весовых категорий не имеют значительного преимущества. Максимально слабые показатели у третьей модели отражают отсутствие значительной статистической взаимосвязи между временем боя и весовой категорией спортсменов, что свидетельствует о сбалансированности правил проведения соревнований. Таким образом мы построили и провели сравнительный анализ трёх линейных регрессионных моделей установив влияние весовой категории на время завершения боя различными способами.

# Список литературы

**Текущий документ не содержит источников.**

1. Ермаков, А. В. Цифровые технологии в спорте как инструмент воспитания / А. В. Ермаков, Е. Н. Скаржинская // Физическая культура и спорт в XXI веке: актуальные проблемы и их решения : Сборник материалов Всероссийской с международным участием научно-практической конференции, Волгоград, 21–22 октября 2020 года. Том 2. – Волгоград: Волгоградская государственная академия физической культуры, 2020. – С. 189-193. – EDN HUVOVT. [↑](#footnote-ref-1)
2. Ермаков, А. В. Сравнительный анализ соревновательной деятельности в армейском рукопашном бое с использованием системы интеллектуального анализа данных ORANGE 3 / А. В. Ермаков, И. О. Яровой // Компьютерный спорт: проблемы и перспективы развития : Материалы Всероссийской научно-практической конференции, Москва, 02 декабря 2021 года. – Москва: Федеральное государственное бюджетное образовательное учреждение высшего образования "Российский государственный университет физической культуры, спорта, молодёжи и туризма (ГЦОЛИФК)", 2021. – С. 45-51. – EDN SLTDSD. [↑](#footnote-ref-2)
3. Ермаков, А. В. Цифровая трансформация физкультурно-спортивного образования (на примере профессий «аналитик данных в ФКИС» и «тренер- аналитик») / А. В. Ермаков // Цифровая трансформация отрасли "физическая культура и спорт": теория, практика, подготовка кадров : Материалы Межрегионального круглого стола, Москва, 22 апреля 2021 года / Под редакцией М.А. Новоселова. – Москва: Федеральное государственное бюджетное образовательное учреждение высшего образования "Российский государственный университет физической культуры, спорта, молодёжи и туризма (ГЦОЛИФК)", 2021. – С. 21-26. – EDN YFUMAM. [↑](#footnote-ref-3)
4. Ермаков, А. В. Физкультурно-спортивное образование - педагогика на основе больших данных / А. В. Ермаков, Е. Н. Скаржинская, Е. А. Сарафанова // Большие данные в образовании: анализ данных как основание принятия управленческих решений : Сборник научных статей I Международной конференции, Москва, 15 октября 2020 года. – Москва: Издательский дом "Дело" РАНХиГС, 2020. – С. 313-323. – EDN WBZPIL. [↑](#footnote-ref-4)
5. Ермаков, А. В. Моделирование темпо-ритмовой структуры соревновательной деятельности в рукопашном бою с использованием систем интеллектуального анализа данных / А. В. Ермаков // Экстремальная деятельность человека. – 2021. – № 3(61). – С. 38-42. – EDN UZJBED. [↑](#footnote-ref-5)
6. Ермаков, А. В. Прогнозирование с использованием методов математического моделирования в спорте высших достижений на примере зимних видов спорта / А. В. Ермаков, П. Е. Мякинченко // Теория и практика физической культуры. – 2021. – № 2. – С. 52-54. – EDN EXFWXQ. [↑](#footnote-ref-6)
7. Gourlay Uvesten F. TEKNISK KARTLÄGGNING AV JUDO: EN ANALYS AV DAMJUNIORERS JUDO: дис. – Linnaeus University, 2022. [↑](#footnote-ref-7)
8. Adam M. A profile of Paweł Nastula’s individual technical-tactical preparation //Archives of Budo Science of Martial Arts and Extreme Sports. – 2013. – Т. 9. – С. 69-75. [↑](#footnote-ref-8)
9. Marek A., Sterkowicz-Przybycień K. The efficiency of tactical and technical actions of the national teams of Japan and Russia at the World Championships in Judo (2013, 2014 and 2015) //Biomedical Human Kinetics. – 2018. – Т. 10. – №. 1. – С. 45. [↑](#footnote-ref-9)
10. Dimitrova N. BIOMECHANICAL ASSESSMENT OF THE PHYSICAL ACTIVITIES OF THE TECHNIQUE UCHI-MATA IN JUDO SPORT //Activities in Physical Education & Sport. – 2017. – Т. 7. – №. 2 [↑](#footnote-ref-10)
11. Kim E. H., Cho D. H., Kwon M. S. A Kinematic Analysis of Uchi-mata (inner thigh reaping throw) by Kumi-kata types in Judo //Korean Journal of Sport Biomechanics. – 2002. – Т. 12. – №. 1. – С. 63-87 [↑](#footnote-ref-11)
12. Minamitani N. et al. Biomechanical properties of judo throwing technique, uchimata, especially for newly developed flamingo technique //ISBS-Conference Proceedings Archive. – 1988 [↑](#footnote-ref-12)
13. Suarez G. R., Davila M. G., Puche P. P. Análisis de factores biomecánicos y comportamentales relacionados con la efectividad del uchi mata, ejecutado por judokas de alto rendimiento //Universidad de Granada. – 2002 [↑](#footnote-ref-13)
14. Kudo K. Judo in Action - Throwing Technique. – Japan Publications Trading Company, 1967.- 128 p. ISBN 10: 0870400746 ISBN 13: 9780870400742. [↑](#footnote-ref-14)
15. Kim E. et al. Biomechanical traits analysis when performing of judo Uchimata by posture and voluntary resistance levels of uke //ISBS-Conference Proceedings Archive. – 2005 [↑](#footnote-ref-15)
16. Kato S., Yamagiwa S. Statistical Extraction Method for Revealing Key Factors from Posture before Initiating Successful Throwing Technique in Judo //Sensors. – 2021. – Т. 21. – №. 17. – С. 5884 [↑](#footnote-ref-16)
17. Yoon H. The Kinetic Analysis of the Lower Extremity Joints when Performing Uchi-mata by Uke's Posture in Judo //Korean Journal of Sport Biomechanics. – 2005. – Т. 15. – №. 2. – С. 167-183 [↑](#footnote-ref-17)
18. Ермаков А.В., Белов А.Г., Скаржинская Е.Н., Новоселов М.А. Программа по визуальному определению амплитуды движений в основных суставах человека (военнослужащего, спортсмена) при выполнении двигательных действий (спортивных и прикладных упражнений) с использованием искусственного интеллекта. Свидетельство о регистрации программы для ЭВМ 2022613196, 01.03.2022. Заявка № 2022612313 от 18.02.2022 [↑](#footnote-ref-18)
19. All Pro MMA Fighters - UFC, Bellator, One <https://www.kaggle.com/datasets/binduvr/pro-mma-fighters> [↑](#footnote-ref-19)
20. UFC/MMA Biggest Dataset With Differentials <https://www.kaggle.com/datasets/danmcinerney/mma-differentials-and-elo> [↑](#footnote-ref-20)
21. Statsmodels <https://www.statsmodels.org/stable/index.html> [↑](#footnote-ref-21)