МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет математики и компьютерных наук**

**Кафедра математических и компьютерных методов**

Допустить к защите

Заведующий кафедрой  
канд. физ.-мат. наук, доцент

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А. В. Лежнев

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2021 г.

Руководитель ООП  
докт. физ.-мат. наук, доцент

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ М. В. Голуб

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2021 г.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА  
(МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)**

**РЕГРЕССИОННЫЕ МОДЕЛИ ДЛЯ ОПРЕДЕЛЕНИЯ ПОВТОРНОГО МАКСИМУМА В ПАУЭРЛИФТИНГЕ**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д. А. Пасько

(подпись)

Направление подготовки 02.04.01 Математика и компьютерные науки

Направленность (профиль) Математическое и компьютерное   
 моделирование

Научный руководитель

докт. физ.-мат. наук,   
доцент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ М. В. Голуб

(подпись)

Нормоконтролер

ст. лаборант \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ И. С. Пшикова

(подпись)

Краснодар

2021

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 3](#_Toc73449002)

[1 Актуальность исследования 5](#_Toc73449003)

[1.1 Зачем предсказывать повторные максимумы? 5](#_Toc73449004)

[1.2 О калькуляторах 6](#_Toc73449005)

[1.3 О данных и ресурсах 8](#_Toc73449006)

[1.4 Отправная точка 9](#_Toc73449007)

[1.5 Гипотеза и допущения 10](#_Toc73449008)

[1.6 Соображения о возможности построения модели 14](#_Toc73449009)

[1.7 План исследования 16](#_Toc73449010)

[2 Описание выборки и разведочный анализ данных 17](#_Toc73449011)

[2.1 О прошедших опрос 17](#_Toc73449012)

[2.2 Взаимодествия переменных 25](#_Toc73449013)

[2.3 Исправление опечаток 36](#_Toc73449014)

[2.4 Удалённые данные 36](#_Toc73449015)

[2.5 Подведение итогов и отбор признаков 42](#_Toc73449016)

[3 Построение моделей 44](#_Toc73449017)

[3.1 Недостатки исходного решения 44](#_Toc73449018)

[3.2 Уточнение коэффициентов 47](#_Toc73449019)

[3.3 Уточнение коэффициентов: зависимость от диапазона 50](#_Toc73449020)

[3.4 Третье поколение моделей 54](#_Toc73449021)

[3.5 Сравнение линейных моделей и подведение итогов 56](#_Toc73449022)

[3.6 Нелинейные модели 60](#_Toc73449023)

[3.6.1 Основные известные модели 60](#_Toc73449024)

[3.6.2 Новые предложения 62](#_Toc73449025)

[3.6.3 Сравнение нелинейных моделей 62](#_Toc73449026)

[ЗАКЛЮЧЕНИЕ 65](#_Toc73449027)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 67](#_Toc73449028)

# ВВЕДЕНИЕ

Как известно, научные исследования, развитие Интернета и компьютерной техники открыли нам возможность создавать и использовать искусственный интеллект в огромном множестве прикладных задач, начиная с простейших моделей предсказания цен на дома и фильтров спама, продолжая распознаванием лиц и очень качественными переводчиками, а заканчивая машинами с автопилотом и генеративными моделями, позволяющими исследовать такие удивительные и, казалось бы, присущие только человеческому мозгу свойства как воображение и индуктивное мышление. Одна из ключевых математических закономерностей, почему это стало возможным, заключается в возможности представления многих задач через пространства низкой размерности, правильные признаки [9]. Иными словами, зачастую огромная часть на самом деле очень сложной закономерности описывается подпространством низкой размерности, коротким набором чисел. Эта идея напрямую реализуется во многих моделях обработки естественного языка, автокодировщиках и даже сжатии изображений вейвлет-преобразованиями; многие алгоритмы классического машинного обучения, в том числе дискриминантный анализ и машины опорных векторов, сводятся к тому, чтобы представить исходные признаки в таком пространстве, в котором решение задачи сильно упростится; нейронные сети хороши в основном тем, что сами генерируют оптимальные признаки, часто непонятные человеку. Сложнейшие задачи социологии, предсказания результатов матчей и, как далее будет показано, задачи предсказания повторных максимумов для человека вполне реально решать с хорошим качеством, если обладать полезными признаками.

Несмотря на всю мощь современных моделей машинного обучения, многие задачи до сих пор лучше решать классическими более простыми методами, привлекая экспертов из конкретной области и более продвинутую статистику. Часто это позволяет быстро сконструировать полезные признаки, избежав переобучения. Именно так мы и поступаем дальше, решая задачу предсказания повторного максимума в пауэрлифтинге – предсказании максимального соревновательного результата в приседании со штангой на плечах, жиме лёжа на горизонтальной скамье, становой тяге. Полученные результаты, как предполагает автор, вполне могут быть обобщены на другие силовые упражнения, где огромная составляющая движения выполняется абсолютной силой, в том числе жимы сидя, жим ногами, другие виды приседаний и тяг.

# 1 Актуальность исследования

## 1.1 Зачем предсказывать повторные максимумы?

При подготовке к соревнованиям по пауэрлифтингу и даже в процессе любительских тренировок полезно (а иногда и необходимо) оценить свои текущие возможности, не делая «проходку», так как «проходка» оказывает сильное воздействие на нервную систему (стресс) и с немалой вероятностью чревата травмами, её нежелательно делать часто и к ней нужно сначала циклически подготавливаться, затем отдыхать (чтобы суметь мобилизовать больше ресурсов и уменьшить риск травм), что занимает много времени и сил – и всё ради того, чтобы просто узнать, какой именно вес ты способен пожать, присесть или потянуть конкретно сейчас. Гораздо разумнее было бы оценить эти величины (*повторные максимумы*, *ПМ*[[1]](#footnote-1)), основываясь на результатах, которые ты показывал совсем недавно в процессе обычных тренировок или которые намного проще продемонстрировать[[2]](#footnote-2) (*многоповторные максимумы*, *МПМ*), то есть используя уже доступную или с меньшими рисками добываемую информацию. Кроме того, *если удастся построить такую модель, способную оценить силу человека, исходя из его последних достижений, можно будет сделать более явной связь между тренировками и реальными результатами, предотвращая перетренированность и временные потери на тренировки по программам, не дающим эффекта*.

Итак. Зачем же нужно предсказывать повторные максимумы? Регрессионные модели, которые требуется построить, помогут:

1) адекватно оценить свои возможности перед соревнованиями и заказать веса, очень близкие к реальному максимуму; кроме того, знать свой ПМ, не делая проходку, очень важно не только для тех, кто занимается пауэрлифтингом (см. [10]);

2) в период подготовки оценивать скорость своего прогресса (или вообще его наличие), предотвращая перетренированность; сравнивать свои результаты с результатами предыдущих циклов, даже если тогда использовалось другое число повторений;

3) более обоснованно планировать программу тренировок: пытаясь чередовать разные упражнения и разные диапазоны повторений, легко сделать программу, которая на самом деле не будет давать значимого эффекта или потребует неподъёмных усилий; чтобы снизить вероятность подобных свойств, программу можно составлять в согласии с дополнительными инструментами, основанными больше чем на опыте нескольких человек

4) обосновать действенность разных методик и выразить различия между ними;

5) более точно замерять силовые качества групп мышц-антагонистов (очень редко для этого вообще возможно измерять 1ПМ у обеих групп, как и МПМ при одном и том же числе повторений);

6) выявить математические закономерности в силовых показателях человека.

Я убеждён, что если какой-то спортсмен действительно имеет цель улучшить свои показатели, он не откажется от дополнительных инструментов, которые способны обеспечить ему преимущество. Конкретно в нашем контексте эти инструменты вычислений называют калькуляторами (повторных максимумов).

## 1.2 О калькуляторах

К сожалению, многие люди отнеслись к этому исследованию очень скептически, поскольку уверены, что здесь имеется слишком сложная зависимость, чтобы её можно было описывать[[3]](#footnote-3); другие утверждали, что знание такой информации принесёт больше вреда, чем пользы, если спортсмен не имеет большого опыта или не тренируется с тренером. Во многом они правы, но, на самом деле, как я считаю, такое мнение обусловлено достаточно большим количеством разных калькуляторов в Интернете, которые дают слишком разные результаты и не ясно на чём основываются (на каких формулах либо откуда полученных).

Примеры:

1) формула из книги Брендона Лилли, с которой началось исследование. За более чем месяц я так и не получил ответа от руководства сайта Juggernaut (<https://www.jtsstrength.com>), откуда взялась эта формула (чуть позже я выяснил, что это округлённая формула Вендлера);

2) калькулятор на (весьма хорошем) сайте Symmetric Strength (<https://symmetricstrength.com>), аналогичная ситуация;

3) формулы Мориса и Райдина [11], которые, как позже будет показано, нерабочие и, скорее всего, получены из слишком маленьких выборок;

4) формулы О Коннора, Бжицки (Бржыки), Лэндера (собранные вместе [12]) и другие, с которыми можно ознакомиться в Википедии [13];

5) также калькуляторы, представленные на сайтах <https://powerliftingrating.ru/repeat_calculator>, <http://evgeniyfit.ru/fitnes-kalkulyatoryi/odnopovtornyiy-maksimum-online/>, <https://body1.ru/kalkulyator-odnopovtornogo-maksimuma-1pm/> и разные клоны одних и тех же калькуляторов.

По-видимому, многие люди уже пытались составить формулы для вычисления повторного максимума, но теперь, помимо самих формул, об этих исследованиях мало что известно.

Цель же этого исследования – предоставить формулу, полученную в согласовании с концепциями машинного обучения, которая будет хорошо работать на многих людях, а не на десяти. Здесь написано, как она получена и почему.

## 1.3 О данных и ресурсах

Для сбора наблюдений был создан опрос в Google Forms[[4]](#footnote-4) на русском языке и его англоязычный аналог[[5]](#footnote-5).

Русскоязычный опрос распространялся через соц. сеть ВКонтакте преимущественно[[6]](#footnote-6) в следующих группах: «Пауэрлифтинг | Тяжелая атлетика»[[7]](#footnote-7), «Я ♥ ПАУЭРЛИФТИНГ»[[8]](#footnote-8), «ПАУЭРЛИФТИНГ И ЖИМ ЛЁЖА 18+»[[9]](#footnote-9), «Твой Тренер»[[10]](#footnote-10); аналогично англоязычный опрос распространялся на Facebook преимущественно здесь: «Powerlifting Motivation»[[11]](#footnote-11), «International Powerlifting League (IPL)»[[12]](#footnote-12), «Powerlifting Motivation Chat»[[13]](#footnote-13), «Powerlifting Memes»[[14]](#footnote-14). Кроме того, несколько наблюдений я собрал лично.

Спустя два месяца от создания опросов данные были выгружены и началась их обработка. Теперь эти данные и всё, связанное с ними (в том числе последняя версия этого документа), хранятся в репозитории по URL: <https://github.com/PasaOpasen/Powerlifting-training-diary-and-articles/tree/master/Estimating%20RM> .

Для обработки данных и создания отчёта использовалась среда RStudio[[15]](#footnote-15) и язык R версии 3.6.3. Скачав некоторые документы, вы сможете воспроизвести все результаты, о которых будет рассказано далее. Для более подробного изучения языка R и практик его использования в статистическом обучении рекомендую ознакомиться с соответствующей литературой (например, [3, 4, 5, 6]).

## 1.4 Отправная точка

В книге «Система тренировок КУБ» [8] на странице 23 (34 в оригинальной версии) приводится формула для определения повторного максимума:

где берётся некоторый рабочий вес и максимально возможное число повторений с ним. В оригинальных обозначениях:

К примеру, если вы можете пожать 100 кг на 5 раз, то из этой формулы следует, что вы сможете пожать 115 на раз и 90 на 9 раз (звучит правдоподобно).

К слову, примерно такая модель с некоторыми поправками на диапазон повторений используется на сайте Symmetric Strength (ссылка приводилась выше). Там максимально возможное количество повторений равно 10.

Лично на мне эта формула хорошо работает и мне захотелось уточнить её для других людей разной комплекции и уровня подготовки. Также интересно то, что здесь имеется, по сути, очень простая линейная модель с двумя *предикторами*: рабочим весом и *взаимодействием* рабочего веса с числом повторений, и нет зависимости от самого упражнения, опыта человека и других характеристик, то есть в перспективе возможно с большой точностью описать наши возможности одной простой формулой, и это будет верно почти для всех людей, невзирая на все различия между ними.

## Гипотеза и допущения

Сила человека в одном повторении зависит от пяти составляющих (факторов) [1, 2]:

1) количество миофибрилл в мышечном волокне. Чем их больше, тем б*о*льшую силу может развить волокно.

2) запасы АТФ и креатинфосфата и уровень ферментации, поддерживающий быстрое преобразование креатинфосфата в АТФ. Если эти запасы малы, спортсмен не сможет поддерживать максимальное усилие нужные несколько секунд. Сам по себе уровень ферментации для силовых нагрузок можно не учитывать, поскольку после нескольких минут хорошей разминки он становится достаточно высоким.

3) соотношение красных и белых мышечных волокон в работающей мышце: чем больше белых волокон, тем б**о**льшую силу способна развить мышца на короткий промежуток времени. Это соотношение является разным для разных мышц и разных людей (определяется генетикой) и, как пока считается, плохо поддаётся коррекции[[16]](#footnote-16).

4) количество вовлечённых в работу волокон. Чем их больше, тем больше сила мышцы. Количество задействованных волокон зависит от *импульса ЦНС* и *порога действия органов Гольджи*: чем сильнее импульс, тем больше волокон может вовлечься в работу, однако импульсы выше некоторого порога подавляются органами Гольджи, чтобы человек не мог сознательно применить силу, способную привести к разрыву сухожилий. Тренировки с субмаксимальными и запредельными весами способны воздействовать как на способность генерировать импульс, так и на порог его подавления[[17]](#footnote-17).

5) естественные рычаги человека и техника выполнения упражнения. У каждого человека есть свои рычаги, дающие ему преимущества в тех или иных типах движений. Кроме того, можно откорректировать технику выполнения упражнения так, что амплитуда движения уменьшится в несколько раз, вдобавок при движении основную работу будут выполнять наиболее сильные пучки мышц. Этот фактор не относится конкретно к силе, но фактически его учёт поможет увеличить демонстрируемые результаты.

Подытожив, можно сказать, что сила спортсмена зависит от генетики (в плане соотношения мышечных волокон, рычагов и т. п.), опыта тренировок (чем больше опыт, тем больше может быть как импульс ЦНС, так и порог его подавления, так и запасы креатинфосфата и т. п.) и базовой конституции (типа телосложения, роста как такового, соотношения роста и функционального веса). Сама техника упражнений учитываться не будет.

*Гипотеза исследования* состоит в том, что одноповторный максимум можно с небольшими ошибками предсказать через многоповторный при помощи модели вида

с, возможно, поправками на телосложение, опыт тренировок, диапазон повторений или некоторые другие факторы.

Здесь (формула (3)) – некоторые числа, – функция, которую ещё придётся подобрать, – ошибка, вносимая неучтёнными факторами. При этом *предполагается, что искомая зависимость очень близка к линейной*, то есть первое слагаемое вносит основной вклад в сумму; это объясняется тем, что интуитивно ясно логическое заключение: из (вес 50 на 6 раз примерно значит 60 на раз) должно следовать с, возможно, небольшими поправками (до 10%) на другие факторы (функция ). С другой стороны, формула должна быть верна независимо от того, указывается вес в килограммах или фунтах (используемых в США).

Такой вид формулы связан с тем, что чуть ли не все описанные ранее факторы, влияющие на силу человека, невозможно измерить, не делая особых анализов крови и т. п., то есть такого рода данные для исследования собрать невозможно. Однако, как будет показано далее, мы всё равно имеем очень полезные признаки, позволяющие делать качественные предсказания.

Поправка на диапазон повторений имеет следующие соображения. Ввиду физиологии в разных диапазонах повторений проявляются разные мышечные способности, не все из которых тесно связаны с целевыми мышечными качествами – абсолютной силой и мощностью. Почти очевидно, что сила человека в одном повторении хорошо коррелирует с силой в двух-пяти повторениях, но то, поднимет ли человек указанный вес на 12 раз или на 20, зависит от его силовой выносливости, которая не вносит большого вклада при работе на одно повторение. Если не вдаваться в подробности, это объясняется тем, что в малом числе повторений основную роль играют запасы креатинфосфата и АТФ, ферментация, напряжение нервного импульса, количество задействованных белых волокон и их сечение, а в большом числе повторений (больше 10-15) существенную роль играет тренированность красных волокон и способность мышц быстро утилизировать продукты метаболических реакций[[18]](#footnote-18) (если утилизация быстрая, спортсмен сможет сделать на 3-5 повторений больше, а если медленная, то даже огромная сила воли не поможет сделать 2-3 лишних повторения из-за чувствительности к падению pH).

Согласно Хетфилду [1], сила развивается исключительно при работе с весами не меньше 80% от ПМ и при этом спортсмены среднего уровня способны выполнить 10-15 повторений с 80% от своего ПМ; то, какой вес человек поднимает на более чем (предположительно) 15 раз, мало говорит о том, какой вес он способен взять на раз.

Кроме того, в статье [2] прослеживается такая идея: разное число повторений требует своего времени на выполнение, а в зависимости от времени работы мышцы включается свой режим энергообмена: примерно через 7 секунд работы запасы креатинфосфата израсходованы больше чем наполовину, алаклатный режим работы (за счёт креатина) завершается через 12 секунд, потом начинается гликолиз (за счёт гликогена), а через 30-60 секунд работы начинается окисление. При этом, если учитывать, что за 30 секунд обычно делается 10 повторений в жиме и 8 в приседе/тяге, можно сделать выводы, что чисто силовые качества проявляют себя (конечно, приблизительно) в первых пяти повторениях, а силовая выносливость заканчивает оказывать влияние на 15-м повторении; всё, что выше 15-20-ти повторений – вообще не имеет отношения к силе[[19]](#footnote-19).

## Соображения о возможности построения модели

К сожалению, планируемая модель априори не сможет идеально точно предсказывать результаты конкретных спортсменов. Существует также множество причин, почему модель, возможно, не выявит реальные закономерности в генеральной совокупности. *Зашумленные данные* и *неполные данные*:

1) опрос проводился почти полностью в Интернете, точность и честность ответов остаются на совести испытуемых. Быть может, кто-то грубо округлил свои данные или назвал данные, которые нельзя считать идентичными (например, разница между выявлением ПМ и МПМ составляла большой срок, отчего уже нельзя говорить, что исследуемые возможности принадлежат одному и тому же *телу* ) либо не совсем понял целей опроса и назвал заниженные результаты (например, тяжёлый разминочный подход на 3 повторения, когда в действительности возможно сделать 5).

2) через опрос невозможно выявить, предположительно, такие имеющие большое значение факторы как естественные рычаги, процент жира в организме (процент веса, который *условно* не имеет вклада в результаты, зато может испортить статистику), особенности техники и качества выполнения движений (каждый по-своему считает, будет ли повторение засчитанным). По этой причине многие люди, которые в рамках опроса имеют идентичные параметры, могут иметь совершенно разные результаты, поскольку при одинаковом росте, весе и возрасте у одного больше функциональной массы, у другого ноги длиннее, один использовал бинты (дающие прибавку 5-15 кг и не считающиеся экипировкой), другой нет, один жал по всем правилам, другой в отбив и т. п.

3) есть множество факторов, оказывающих большое влияние, трудно измеримых и при этом варьирующихся у одного и того же человека. Например, при одних и тех же мышцах человек может поднять один и тот же вес на 3 или 5 раз в зависимости от уровня креатинфосфата и психологического настроя; при следующей тренировке эти параметры могут быть уже значимо другими, отчего правила, хорошо работавшие недавно, перестанут хорошо работать сейчас.

В то же время имеются доводы в пользу модели:

1) ПМ предсказывается не только на основе роста или веса, но и на основе многоповторного максимума, то есть не многим меньшего веса и числа повторений с ним. Нельзя отрицать, что МПМ, всё же, содержит немалую информацию о ПМ. Как было замечено авторами книги «Анализ данных в науке и технике» [9], *«во многих естественно возникающих системах в данных присутствуют преобладающие паттерны, характеризуемые аттрактором или многообразием низкой размерности»*, то есть часто огромную долю на самом деле сложной закономерности описывает подпространство низкой размерности, несколько очень информативных признаков; в нашем случае МПМ и диапазон повторений — как раз такие признаки.

2) существует модель Вендлера (и схожие подели), которая приемлемо работает для многих спортсменов. Сам факт наличия такой модели говорит о том, что ПМ можно описать через МПМ в какой-то степени или хотя бы для какой-то группы людей.

Можно долго рассуждать о том, будет ли работать модель или нет и почему, но для того и проводится исследование, чтобы найти математически обоснованный ответ.

## 1.7 План исследования

Сперва будет произведён разведочных анализ собранных данных с целью выявить в них некоторые закономерности, подтвердить или опровергнуть некоторые простые гипотезы, обнаружить выбросы, принять решение о включении тех или иных факторов в модель.

После этого будет построена серия моделей, среди которых будет выбрана лучшая. По правилам машинного обучения лучшей моделью считается та, что даёт наименьшую ошибку при перекрёстной проверке и (согласно принципу бритвы Оккама) наиболее проста в объяснении (имеет хорошее значение информационного критерия). Параллельно будем проверять достоверность уже известных моделей.

В конце я сделаю выводы и опишу простейшие возможности применения полученной модели.

# 2 Описание выборки и разведочный анализ данных

## 2.1 О прошедших опрос

С помощью опроса было получено всего 174 достоверных наблюдения (исключая явные аномалии), принадлежащие предположительно 157 людям[[20]](#footnote-20); они содержатся в файле по ссылке URL: [https://github.com/PasaOpasen/Powerlifting-training-diary-and-articles /blob/ master/](https://github.com/PasaOpasen/Powerlifting-training-diary-and-articles%20/blob/%20master/)Estimating%20RM/data(rus).tsv.

Все наблюдения содержат информацию о нескольких переменных. После некоторых преобразований над этими переменными получаем следующий набор предикторов:

– *RM* – собственно повторный максимум,

– *MRM* – многоповторный максимум,

– *Count* – количество повторений для многоповторного максимума,

– *Action* – движение для которого верны измерения (Жим, Тяга, Присед – самые базовые упражнения),

– *Sex* – пол испытуемого,

– *Experience* – группа опыта (До двух лет, 2-3 года, 4-5 лет, 6-10 лет, больше 10 лет),

– *Age* – возраст,

– *Weight* – собственный вес,

– *Height* – рост,

– *BodyType* – тип телосложения (Эктоморф, Мезоморф, Эндоморф). Типы телосложения в целом различаются скоростью обмена веществ и строением скелета, рычагами (хотя различия в строении скелета во многом являются следствием различий в обмене веществ),

– *CountGroup* – группа по диапазону повторений (2-3, 4-6, 7-10, 11-20, >20),

– *AgeGroup* – возрастная группа (<20, 20-27, 28-35, >35),

– *Index* – индекс массы тела[[21]](#footnote-21), *ИМТ*,

– *IndexGroup* – группа по индексу массы тела (выраженный дефицит, дефицит, норма, избыток, ожирение1, ожирение2, ожирение3 в соответствии с рекомендациями ВОЗ).

Основные статистики по данным переменным приведены на рисунке 1.

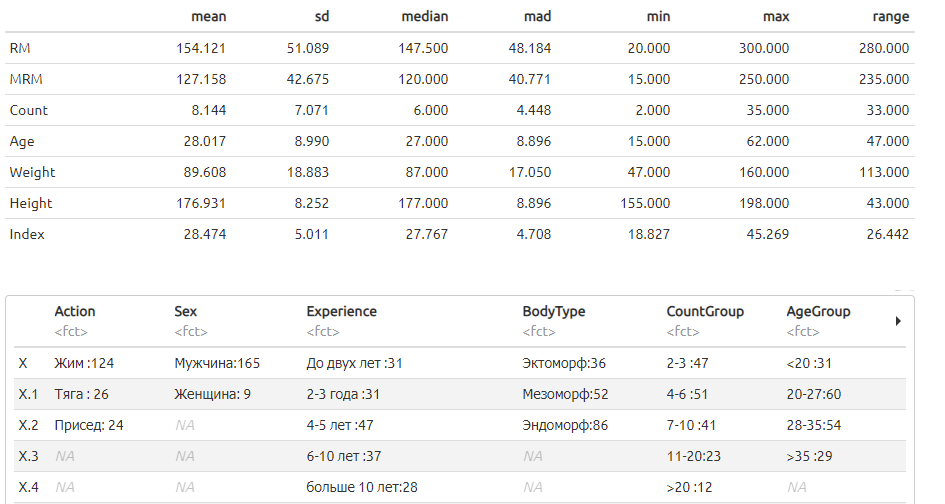


Рисунок 1 – Основные статистики по выборке

Сделаем некоторые выводы о выборке. Почти все испытуемые – мужчины (рисунок 2), поэтому *результаты, которые будут получены, не следует обобщать на женщин*, даже если для женщин из выборки предсказания получаются хорошие; сила, на самом деле, не зависит от пола, но она зависит от поперечного сечения мышц, которое у женщин, как правило, меньше, однако по имеющимся данным выявить эту разницу невозможно. Возможно даже, что позднее придётся удалить принадлежащие женщинам наблюдения из выборки, если окажется, что те сильно выделяются.

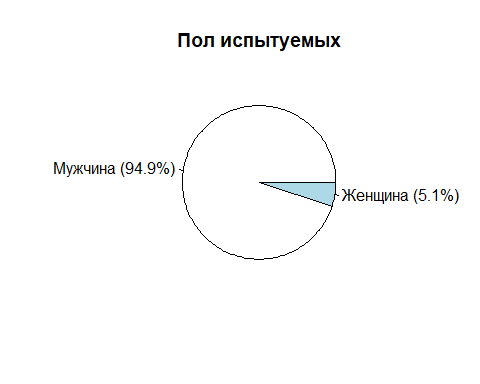


Рисунок 2 – Пол испытуемых

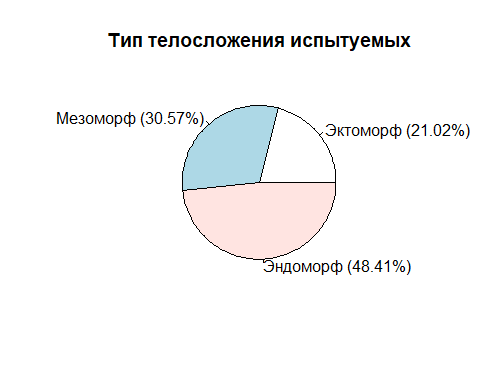
Среди испытуемый почти половину составляли эндоморфы (рисунок 3) 

Рисунок 3 – Тип телосложения испытуемых

и тест хи-квадрат говорит, что это отличие статистически значимо (рисунок 4), то есть *среди пауэрлифтеров эндоморфы встречаются чаще эктоморфов или мезоморфов*:

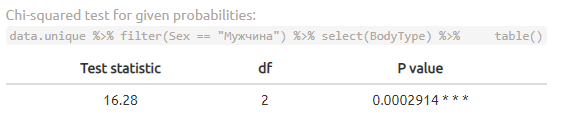


Рисунок 4 – Результаты теста хм-квадрат

Больше половины наблюдений относятся к жиму лёжа (рисунок 5). Я думаю, это связано с тем, что большинство спортсменов просто предпочитают это упражнение двум другим, вдобавок на жиме лёжа относительно проще замерять МПМ и, скорее всего, многие из ответивших специализировались именно на жиме лёжа (жимовики).

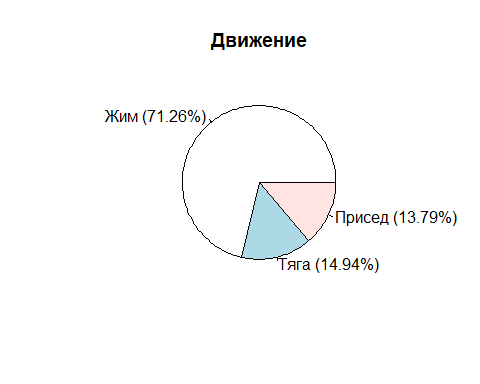


Рисунок 5 – Тип движения в выборке

Распределение по другим факторам иллюстрируется на рисунках 6-11.

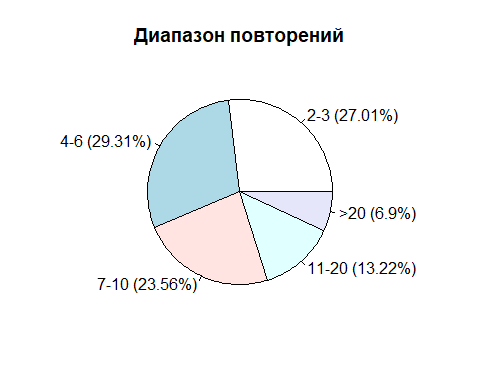


Рисунок 6 – Диапазон повторений в выборке

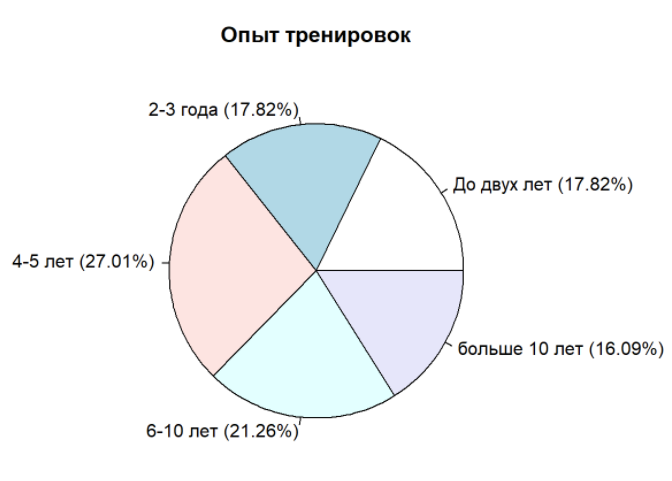


Рисунок 7 – Опыт тренировок в выборке

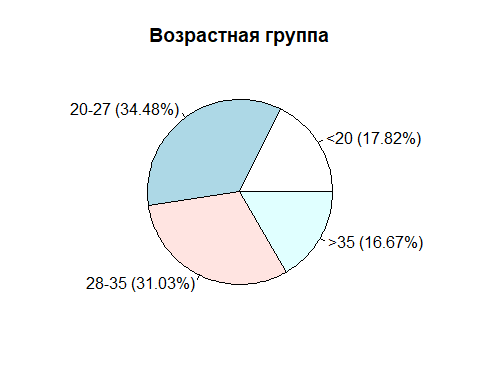


Рисунок 8 – Распределение возрастной группы по выборке

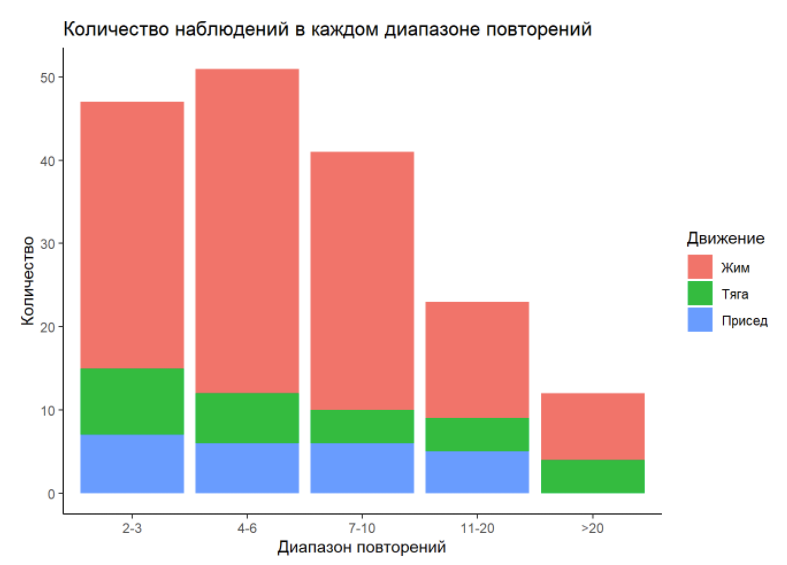


Рисунок 9 – Количество наблюдений в каждом диапазоне повторений

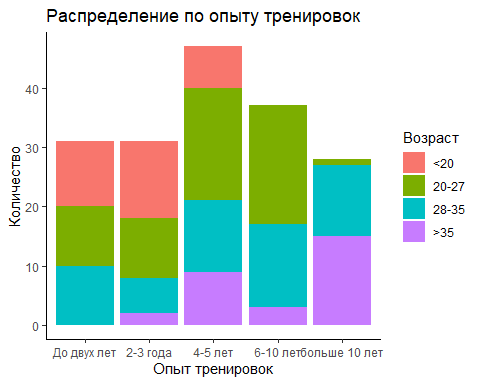


Рисунок 10 – Распределение по опыту тренировок

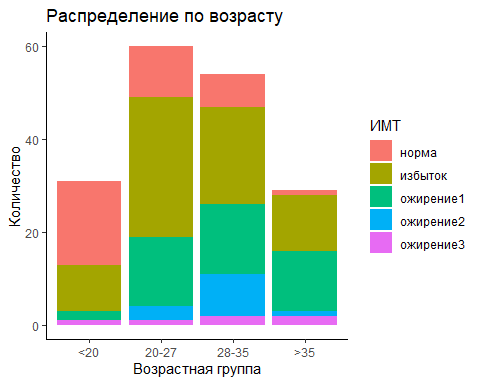


Рисунок 11 – Распределение по возрасту

Распределение по индексу массы тела отображено на рисунках 12-13.



Рисунок 12 – Количество наблюдений по ИМТ

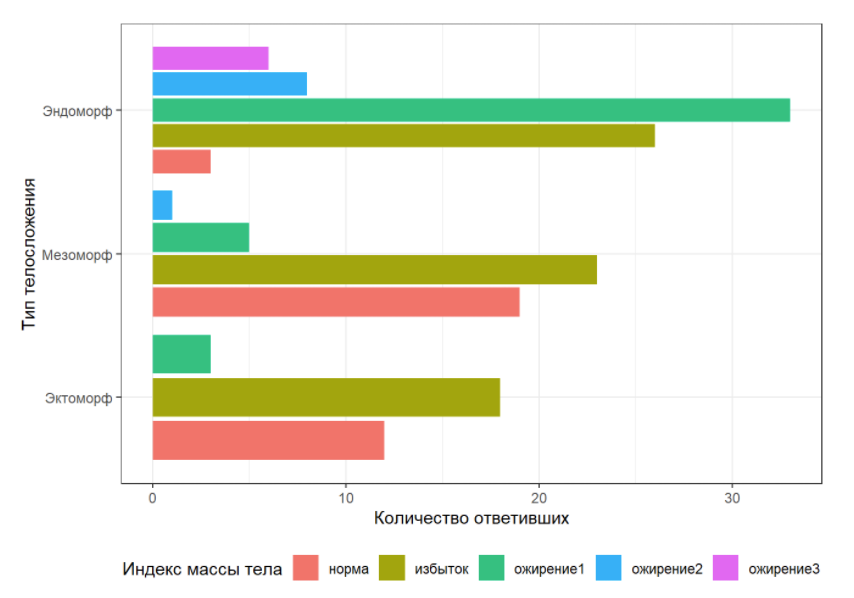


Рисунок 13 – Распределение по ИМТ

Видно, что среди ответивших нет людей с дефицитом веса и немалая часть имеет нормальный вес, немалая – «избыток» (предположительно за счёт мышечной массы), но так называемое «ожирение1» (с большой вероятностью связанное с жировыми отложениями) у эндоморфов встречается чаще, чем в других телосложениях. Это подтверждает тест пропорций (рисунок14).

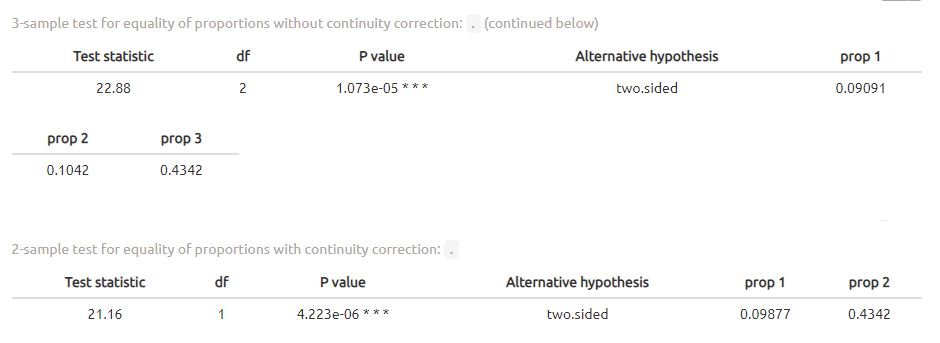


Рисунок 14 – Результаты теста пропорций

Аналогично дисперсионный анализ показывает, что эндоморфы имеют больший ИМТ (рисунок 15).

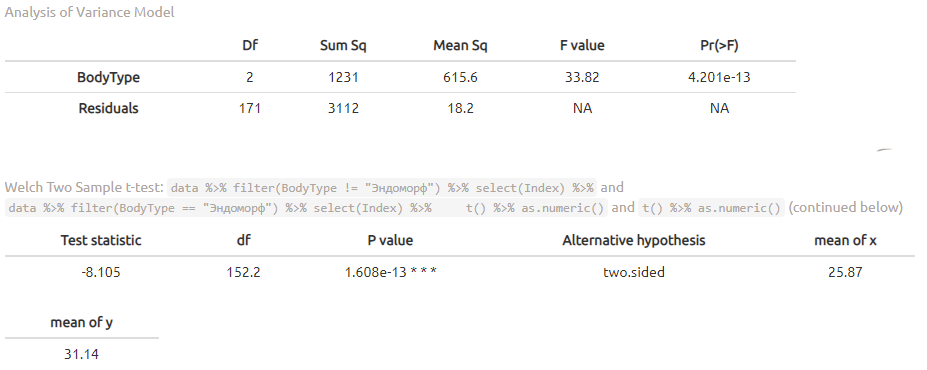


Рисунок 15 – Результаты дисперсионного анализа относительно ИМТ в разных типах телосложений

## 2.2 Взаимодествия переменных

Корреляции количественных переменных приведены на рисунке 16.

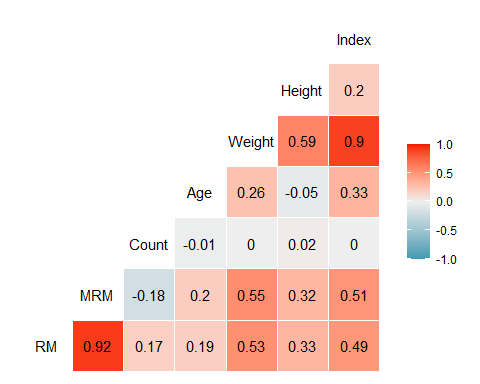


Рисунок 16 – Корреляции количественных переменных

Зависимость повторного максимума от индекса массы тела изображена на рисунках 17-18.

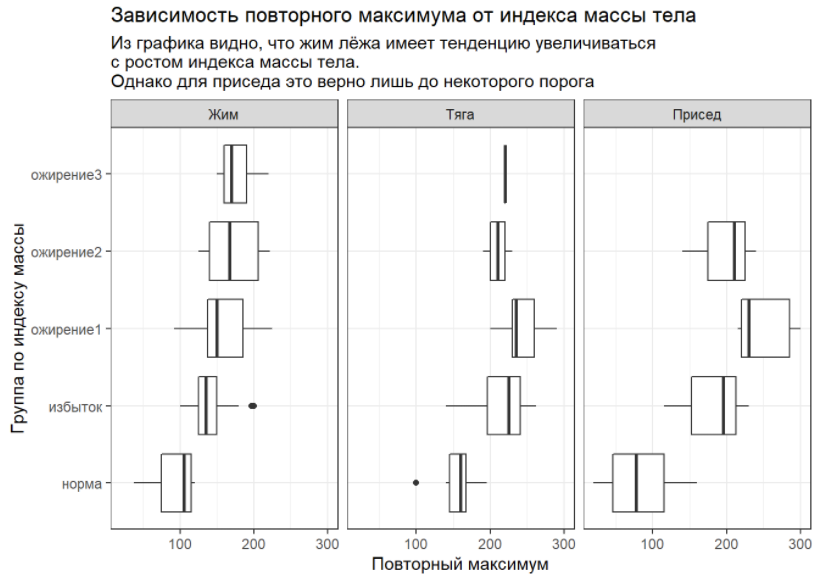


Рисунок 17 – Зависимость повторного максимума от индекса массы тела

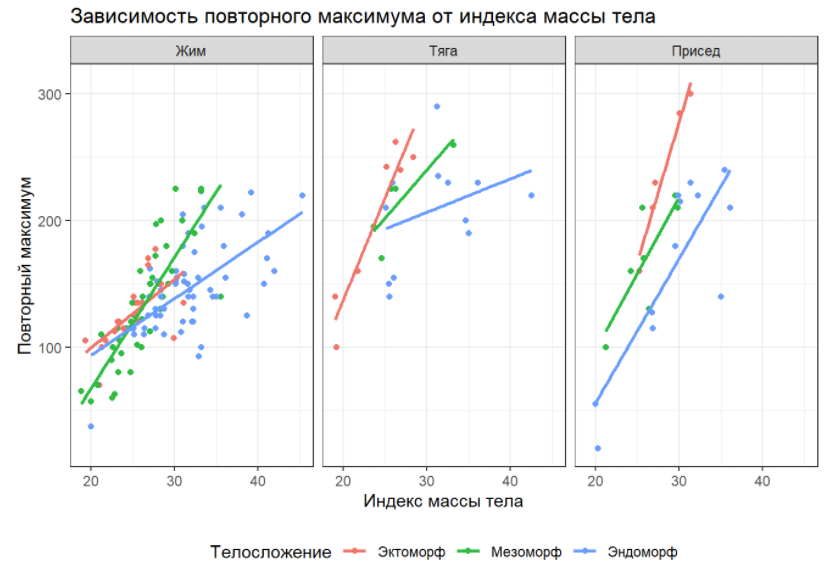


Рисунок 18 – Зависимость повторного максимума от индекса массы тела (разбитое по движениям)

Оценка коэффициента простой регрессионной модели в зависимости от числа повторений приведена на рисунке 19.

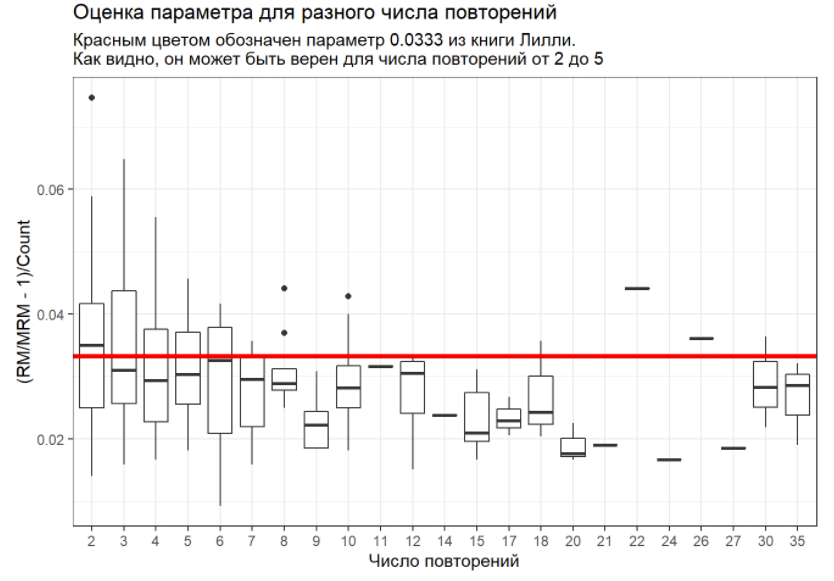


Рисунок 19 – Оценка параметров для разного числа повторений

Из рисунка 20 может следовать, что начиная с 8-9 повторений зависимость между МПМ и ПМ ослабевает.

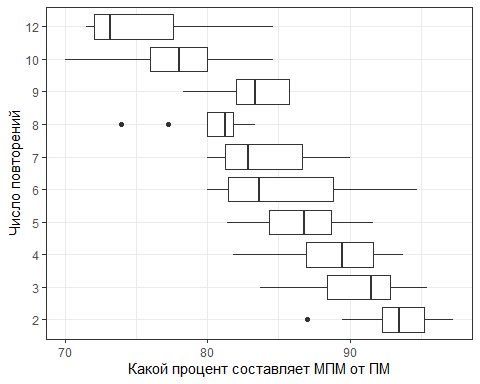


Рисунок 20 – Доли МПМ от ПМ

Либо здесь имеется проблема с данными, так как на 9-м повторении корреляция резко уменьшается (мало наблюдений), рисунок 21.



Рисунок 21 – Корреляции по числу повторений

Есть ли разница в этих процентах для разных движений или телосложений? Дисперсионный анализ показывает, что нет, значимых различий не обнаружено (все p-значения больше 0.05), рисунок 22.



Рисунок 22 – Результаты дисперсионного анализа по числу повторений

Применим тест Стьюдента для определения предсказываемых процентов и доверительных интервалов для них (рисунок 23).



Рисунок 23 – Результаты теста Стьюдента по числу повторений

В целом, эти данные согласуются с тем, что используются *National Strength and Conditioning Association (NSCA)* [14] (рисунок 24).

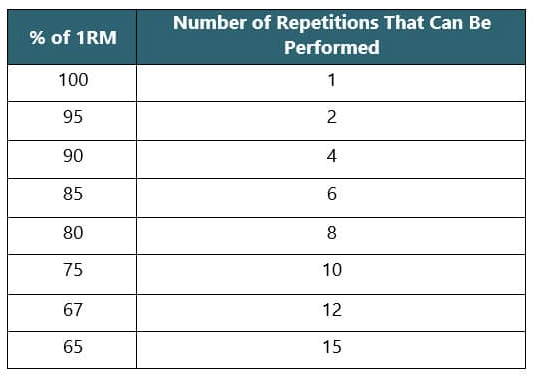


Рисунок 24 – Процентовки от NSCA

Хотя есть и отличия: вес на 2 повторения скорее равен 94% от максимума, а не 95%, а вес на 12 повторений – это скорее 70% от максимума, а не 67%.

Теперь посмотрим на зависимость повторного максимума от многоповторного (рисунок 25).

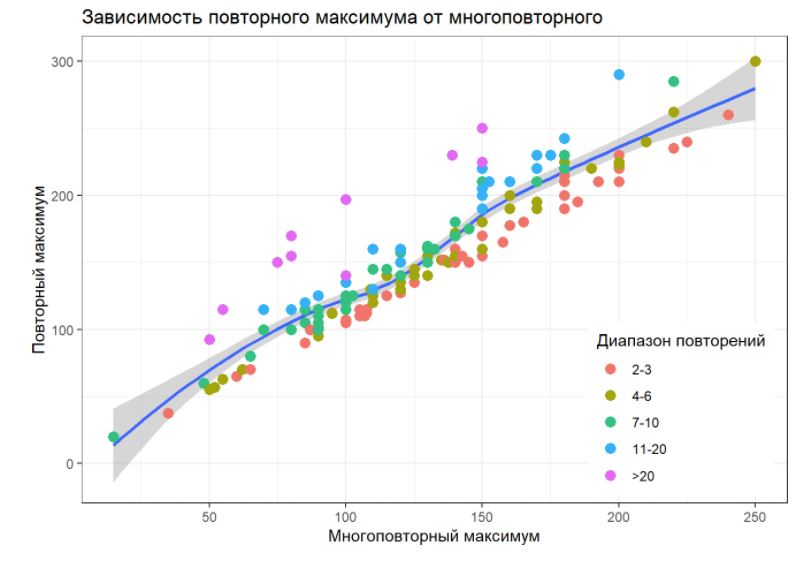


Рисунок 25 – Точечная диаграмма зависимости ПМ от МПМ

Очевидно, что здесь будет близкая к линейной зависимость (учитывая высокую корреляцию).

Оказывается, если предварительно разбить наблюдения на группы по повторениями, линейность станет намного более выраженной для числа повторений не больше 10 (для большего числа повторений не прослеживается явной линейности либо из-за недостатка данных, либо по физиологическим причинам, озвученным ранее; для диапазона 11-20 возможна квадратичная зависимость, но мало данных, чтобы что-то подтвердить), см. рисунок 26.

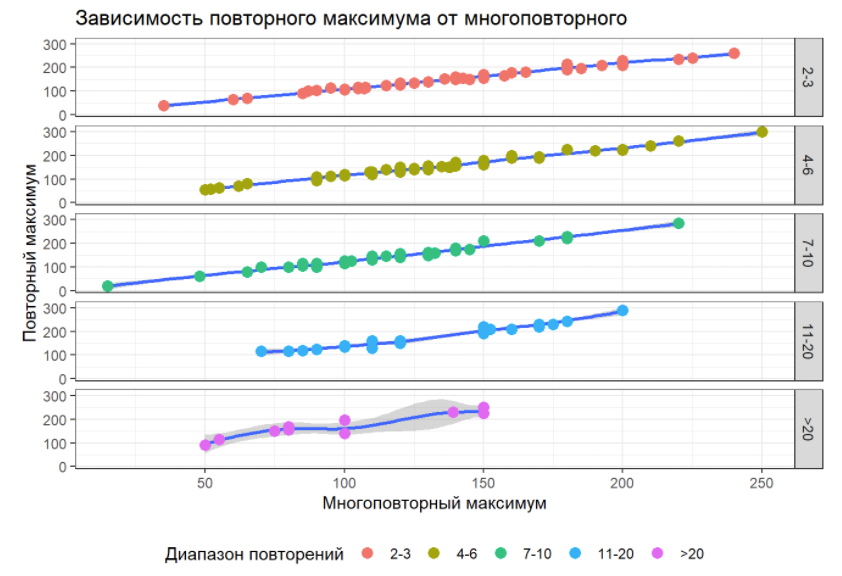


Рисунок 26 – Зависимость ПМ от МПМ, по диапазонам повторений

Сказанное выше значит, что для каждого диапазона повторений нужен отдельный анализ.

Есть ли зависимость между отношением в зависимости от опыта тренировок и других факторов? Дисперсионный анализ показывает, что нет (рисунок 27).

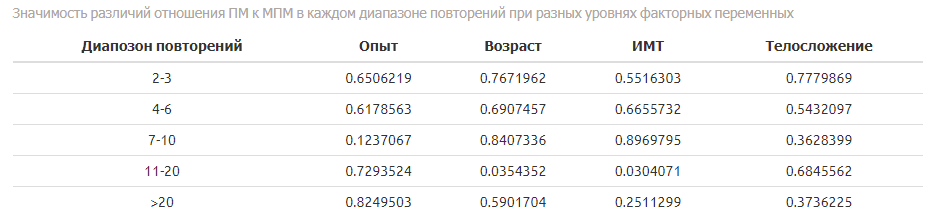


Рисунок 27 – Дисперсионный анализ по диапазонам повторений

Единственное: обнаружилась разница в зависимости от возраста и индекса массы тела для диапазона повторений 11-20 (рисунок 28). Возможно, это связано с тем, что в целом с увеличением возраста увеличивается уровень подготовки, отчего на более чем 10 раз удаётся поднимать больший процент от максимального веса.

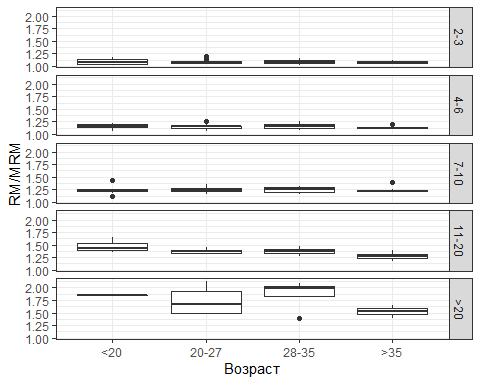


Рисунок 28 – Зависимость от возраста в разных диапазонах повторений

Кроме того, возможна разница между самими упражнениями (движениями), рисунок 29.

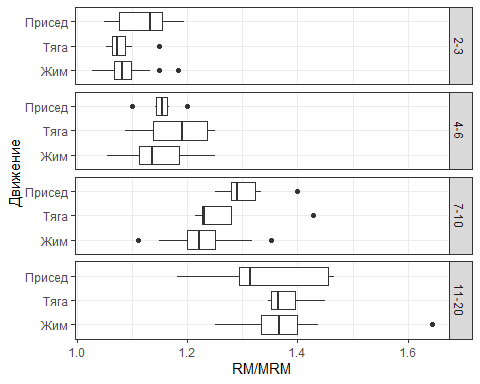


Рисунок 29 – Отношение ПМ и МПМ

Дополнительные закономерности изображены на рисунках 30-32.

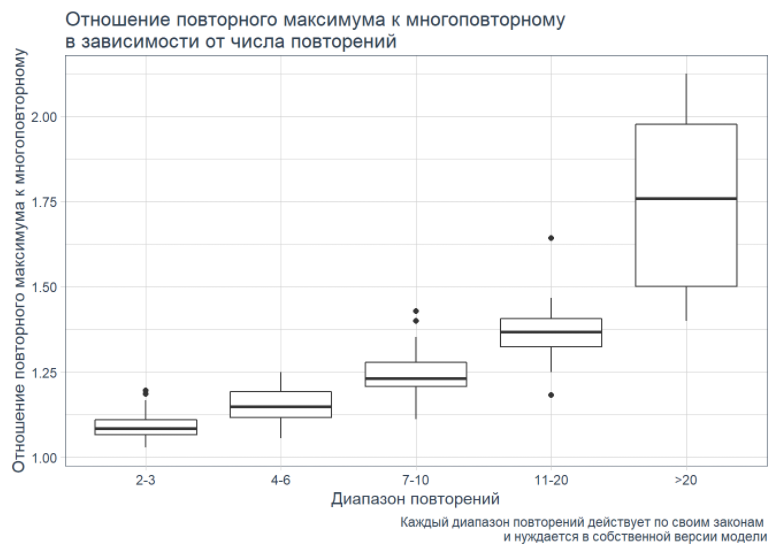


Рисунок 30 – Отношение ПМ к МПМ в зависимости от диапазона и числа повторений

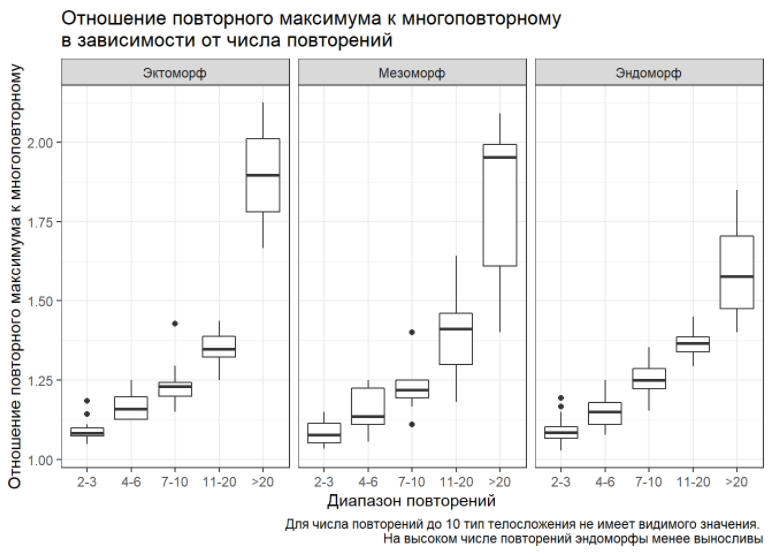


Рисунок 31 – То же отношение, разбитое по типу телосложения

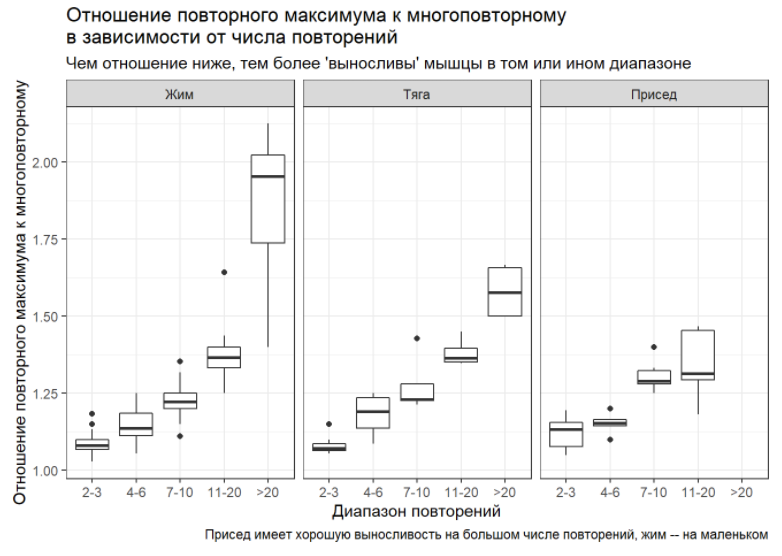


Рисунок 32 – То же отношение, разбитое по движению

## 2.3 Исправление опечаток

Часто при сборе данных возникают *ненормальные* наблюдения, полученные из-за опечаток или недопониманий. В этот раз таких наблюдений накопилось около 10%. Если поразмыслить над их причинностью, то чаще всего испытуемый забывал поставить цифру «1» перед числом, в результате вместо роста 175 получался рост 75 и т. п.

В то же время могло быть и так, что испытуемый неверно понял суть исследования и в анкете описывал просто подход, который был выполнен, а не подход для многоповторного максимума (то есть на пределе усилий), либо при проходке использовал экипировку или другую технику упражнения, либо, опять же, опечатался; в итоге жим 170х5 оказывался равным 220, хотя ясно, что тут должны стоять 170х5 и 200 либо 170х15 и 220.

С такими *выбросами* можно сделать только две вещи:

1) исправить их, опираясь на логику либо переспросить того, кому это наблюдение принадлежит[[22]](#footnote-22);

2) удалить из выборки.

В итоге некоторые данные были исправлены на более вероятный вариант, а явные непонятные аномалии были удалены из выборки. В файле с исходными данными[[23]](#footnote-23) содержатся комментарии, где и что я исправил и почему.

## Удалённые данные

Перед построением модели из обучающей выборки заранее были исключены следующие наблюдения:

1) некоторые выбросы, в том числе несколько наблюдений, которые изначально не вписывались в модель и по логике очень сомнительны; речь идёт о наблюдениях типа , хотя интуитивно понятно, что 140 – это как минимум 120 на 5; возможно, подход 120 на 3 был сделан в достаточно отдалённый от проходки период либо во время проходки в качестве разминочного, то есть его нельзя считать многоповторным максимумом, эквивалентным повторному максимуму; конечно, здесь может играть роль какой-то особый случай[[24]](#footnote-24), который является выбросом (потому должен быть исключён) либо не может быть выявлен по тому набору признаков, которые доступны в нашем случае;

2) наблюдения с числом повторений больше 20, так как сама идея предсказать 1ПМ по, например, 30ПМ – абсурдна, поскольку здесь имеют место в корне разные нагрузки, задействуются разные типы мышечных волокон и имеет место разный энергетический обмен; по этой причине использование таких данных при создании моделей приводит к искажениям, а для самих этих наблюдений модель предсказывает лишь какую-то нижнюю границу 1ПМ, в чём нет смысла[[25]](#footnote-25); вдобавок, для таких предсказаний сложнее определить аномальные наблюдения в смысле предыдущего пункта; кроме того, даже на графиках видно, что диапазон выше 20 очень сильно отличается от остальных диапазонов.

Как итог, теперь для данных верны следующие статистики (рисунки 33-40).

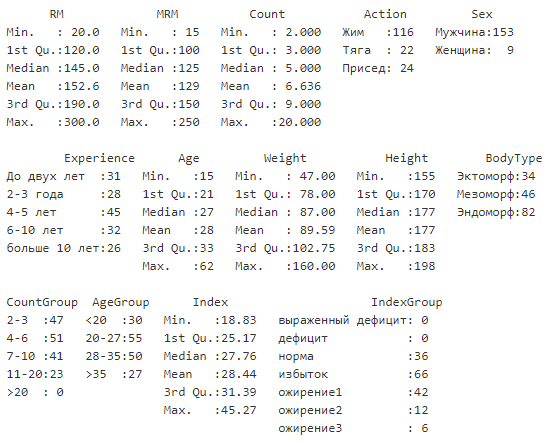


Рисунок 33 – Статистики по измененной выборке

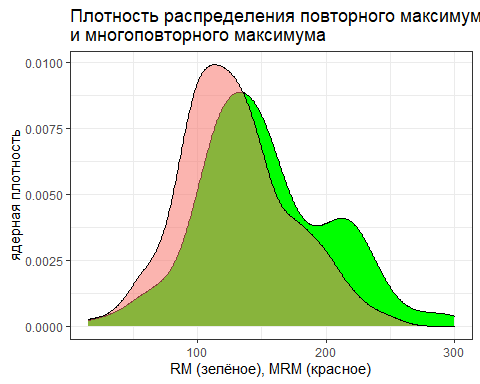


Рисунок 34 – Плотность распределений ПМ и МПМ

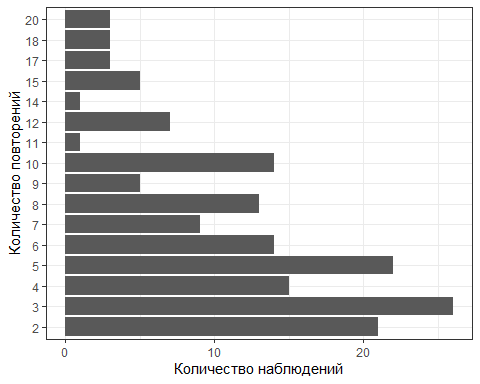


Рисунок 35 – Число повторений в изменённой выборке

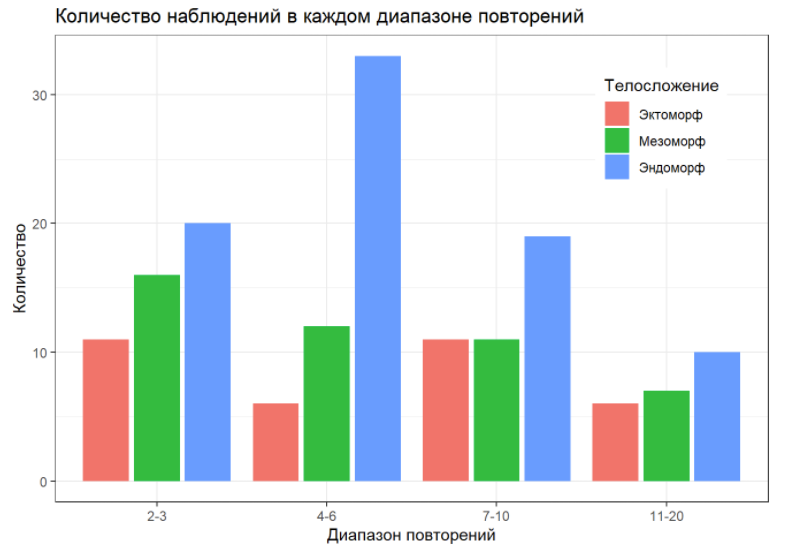


Рисунок 36 – Количество наблюдений в каждом диапазоне повторений



Рисунок 37 – Количество наблюдений в каждой категории по ИМТ

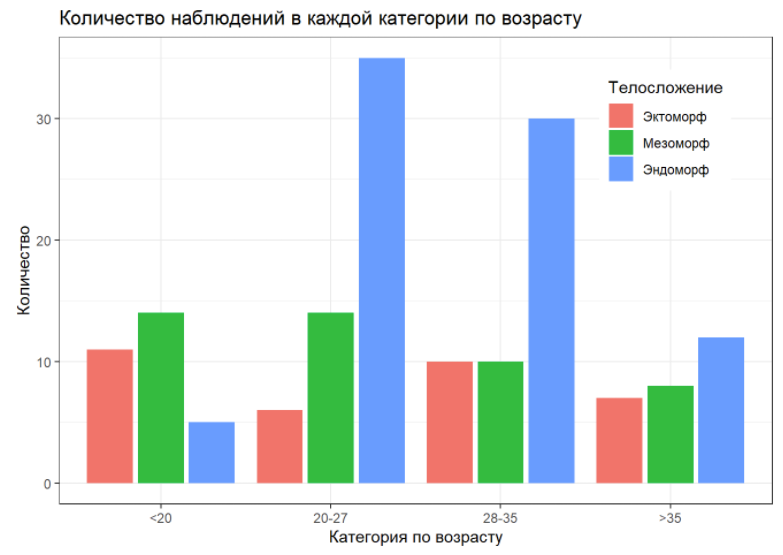


Рисунок 38 – Количество наблюдений в категориях по возрасту

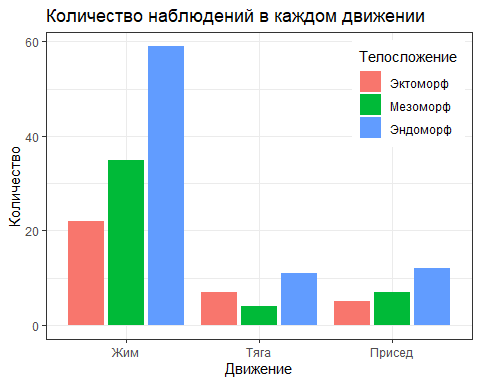


Рисунок 39 – Количество наблюдений в каждом движении

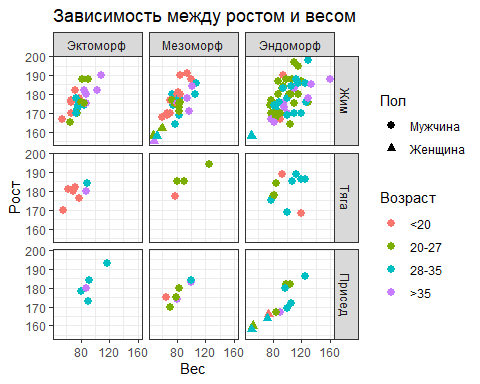


Рисунок 40 – Зависимость между ростом и весом

## Подведение итогов и отбор признаков

Перечислим основные идеи об особенностях данных, сказанные ранее:

1) любые полученные результаты не следует обобщать на женщин, так как в опросе они почти не участвовали. Возможно, ради точности вычислений даже придётся убрать женщин из выборки;

2) эндоморфы среди пауэрлифтеров встречаются чаще мезоморфов или эктоморфов. При этом у эндоморфов значимо чаще встречается «ожирение первой стадии», чем у представителей других телосложений;

3) результаты в приседе растут с увеличением индекса массы тела лишь до некоторого порога. Увеличение индекса массы тела выше “ожирения первой стадии” не будет полезным;

4) сильная корреляция между индексом массы тела и повторным максимумом наблюдается у мезофорфов – в жиме, у эктоморфов – в тяге, у эндоморфов – в приседе;

5) не обнаружено значимой разницы значений для разных телосложений, движений, групп по индексу массы, опыта, возраста для любого конкретного числа повторений или диапазона повторений (но при этом обнаружилась разница между движениями для диапазонов повторений до 10); процентовки от *National Strength and Conditioning Association (NSCA)* с небольшими погрешностями верны для всех спортсменов и всех движений;

6) разбиение наблюдений на группы по диапазону повторений (который сделан согласовано с физиологическими соображениями) должно значительно повысить точность модели, причем для числа повторений до 10-20 эта модель крайне близка к линейной;

7) наблюдения свыше 20 повторений не будут учитываться, так как их мало и этот диапазон слабо коррелирует с абсолютной силой;

8) при более чем 10 повторений опытные спортсмены способны поднимать более близкие к максимальным веса, чем неопытные.

Соображения о предикторах для модели:

1) (повторный максимум) обязательно зависит как от (многоповторный максимум) и (числа повторений), так и от – группы по числу повторений;

2) может зависеть также от (типа телосложения) и (упражнения) либо от их бинарных модификаций (например, если людей разделить не на 3 телосложения, а на эндоморфов и не эндоморфов); возможно, стоит также попробовать учесть индекс массы тела или собственно вес, имеющие корреляцию с не меньше 0.5.

Дальнейшие действия:

1) просмотреть, как работает модель Вендлера, и на её примере ознакомиться с шаблоном по оценке качества, который будет использоваться для оценки следующих моделей.

2) построить разные модели для оценки конкретно и выбрать ту, которая более проста и даёт меньшую ошибку кросс-валидации.

3) сравнить выбранные модели и определить лучшую из них.

# 3 Построение моделей

Вычислительные возможности позволяют создать и проверить огромное количество моделей. Цель исследователя – предложить несколько вариантов и выбрать среди них лучший. Кроме этого, нужно посмотреть, как работает модель на разных группах данных, чтобы обнаружить выбросы или обнаружить, что на такой-то группе модель вообще не может работать, из-за чего такую группу придётся исключить (эта работа уже проделана при самом исследовании, чтобы не нагружать документ).

Итак, в модели представляют интерес следующие показатели:

– ошибки в разных группах данных;

– выбросы и влиятельные наблюдения;

– статистическая значимость модели и её составляющих;

– оценки кросс-валидации.

## 3.1 Недостатки исходного решения

Применив формулу ко всем наблюдениям, получим оценки (Fact), которые будут отличаться от истинного (Target). Разница между этими величинами – остатки. Для модели Вендлера остатки приведены на рисунках 41-43.

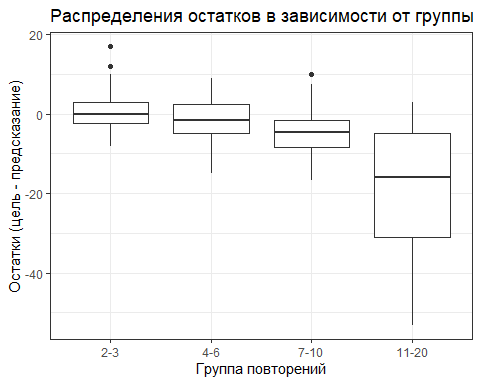


Рисунок 41 – Распределения остатков в зависимости от группы

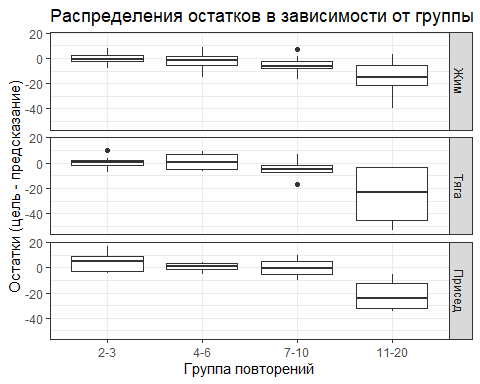


Рисунок 42 – То же самое, разбитое по движению

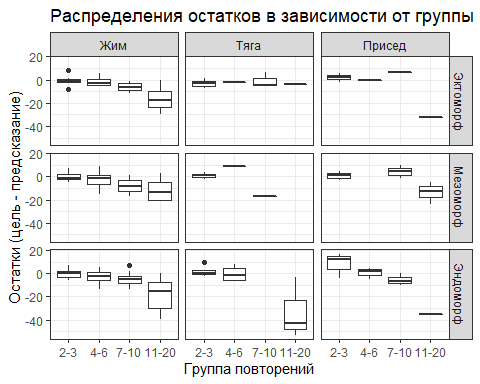


Рисунок 43 – То же самое, разбитое по телосложению

Здесь такая логика: чем лучше модель, тем «ящики» тоньше и тем их середины ближе к 0.

В данном случае видно, что модель Вендлера хорошо работает почти на всех данных из диапазона 2-3 и на немалой части данных из диапазона 4-6. Для большего числа повторений модель даёт завышенные оценки.

Если посмотреть на численные ошибки увидим, что модель заметно ошибается в 30% случаев и в среднем на почти 5%. Первые формулы Мориса и Райдина (ссылка приведена выше) показывают худший результат (40% случаев с большими ошибками)

На самом деле это две формулы вида

где зависят от того, к какой части тела относится упражнение. Сама идея сделать логарифмическое преобразование используется нередко, но здесь это имеет негативный эффект: равен далеко не 0.99, как сказано на сайте (где я нашёл эти формулы), реальный составляет 0.75.

Скорее всего, формулы Мориса и Райдина подгонялись на маленькой выборке или на выборке, в которой почти все данные принадлежали нескольким людям.

## 3.2 Уточнение коэффициентов

Теперь возьмём модель вида и подберём коэффициенты лучшим образом. В итоге получим результаты, изображенные на рисунках 44-49.

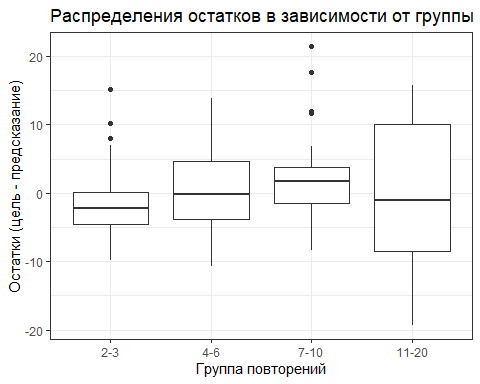


Рисунок 44 – Распределение остатков в зависимости от группы

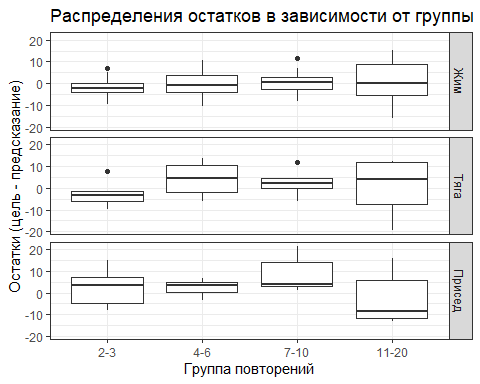


Рисунок 45 – То же самое, разбитое по движению

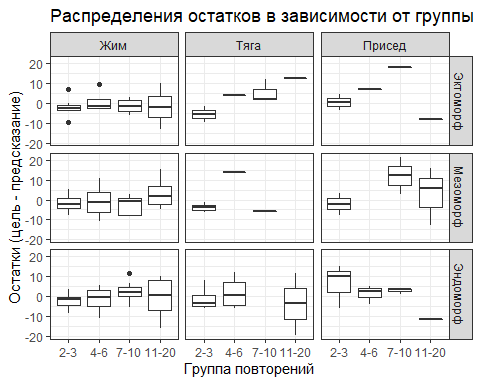


Рисунок 46 – То же самое, разбитое по телосложению

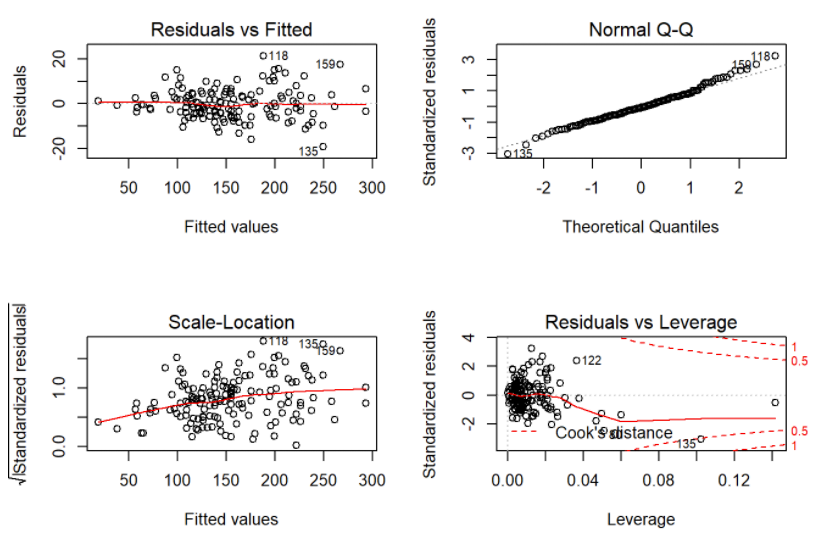


Рисунок 47 – Статистики по качеству модели

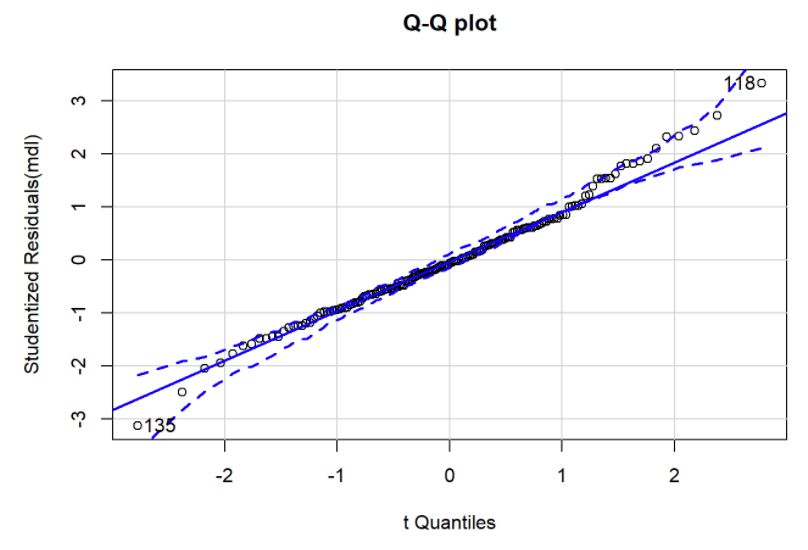


Рисунок 48 – График квантиль-квантиль

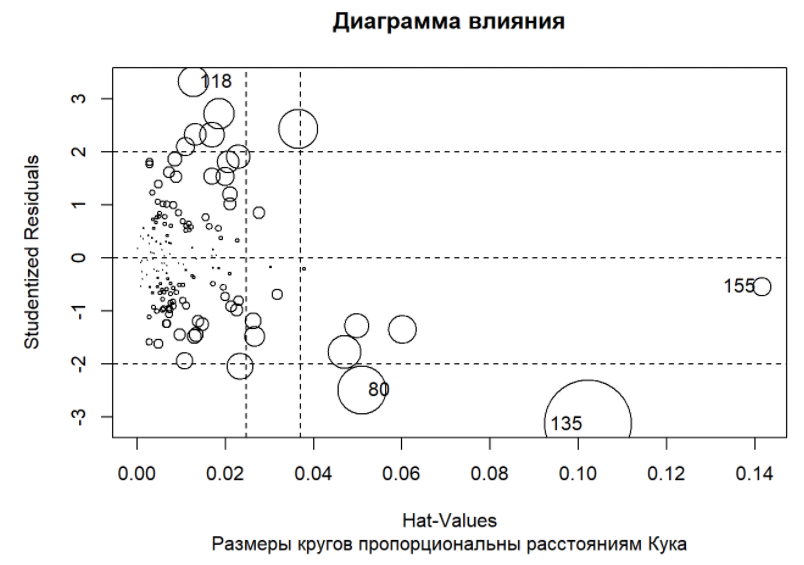


Рисунок 49 – Диаграмма влияния

Далее приведена вся необходимая информация о качестве модели (назовём эту модель ):

1) на выборочных данных модель ошибается максимум на 21кг и 13%;

2) в среднем модель ошибается на 3%;

3) в 75% случаев ошибка не превышала 5%;

4) наибольший разброс ошибок приходится на диапазон 11-20;

5) модель статистически значима и удовлетворяет всем нужным требованиям, кроме гетероскедастичности;

6) немало больших ошибок приходится на диапазон 11-20 повторений, содержащий меньше всего наблюдений.

## 3.3 Уточнение коэффициентов: зависимость от диапазона

Теперь сделаем поправку для коэффициентов в зависимости от факторных переменных.

Путём подбора удалось найти две схожие модели (назовём их и ), немного отличающиеся оценками кросс-валидации на разных диапазонах. Поскольку на диапазоне повторений 11-20 по-прежнему сохранялись сильные ошибки, было принято решение удалить этот диапазон (также это можно обосновать тем, что на этом диапазоне меньше данных и сложнее отлавливать выбросы, да и физиологически он не так тесно связан с силой, как другие, о чём уже было сказано).

Далее приведён анализ для модели , в которой и и имеют поправку на диапазон повторений. Результаты представлены на рисунках 50-54.

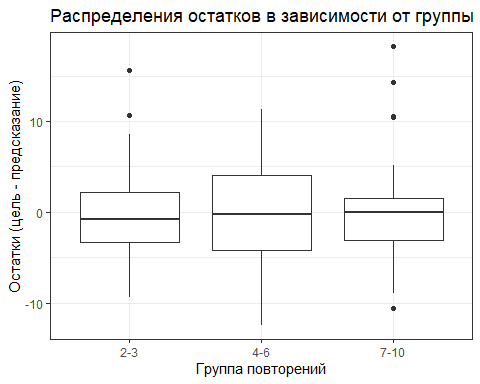


Рисунок 50 – Распределение остатков в зависимости от группы

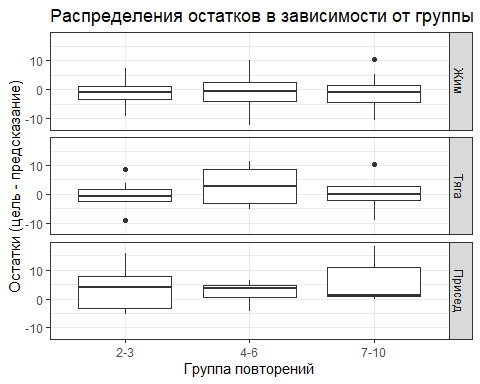


Рисунок 51 – То же самое, разбитое по движению

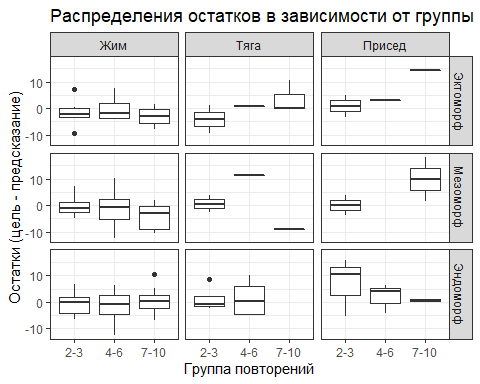


Рисунок 52 – То же самое, разбитое по телосложению

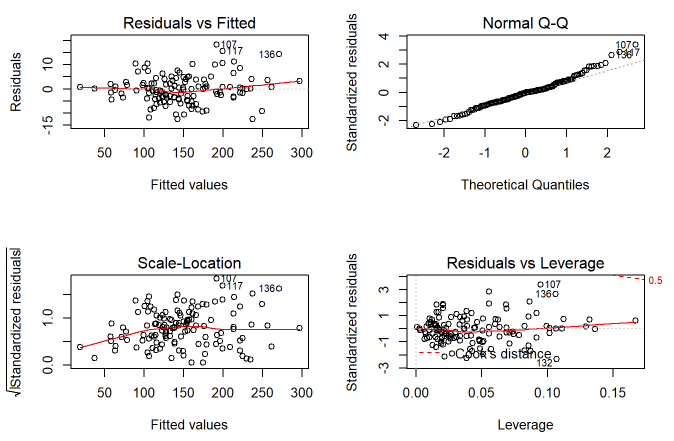


Рисунок 53 – Статистики по качеству модели

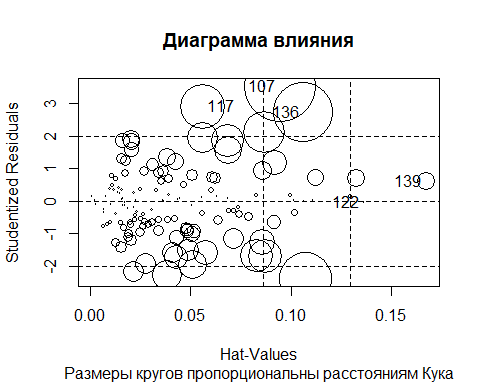


Рисунок 54 – Диаграмма влияния

Основные свойства:

1) на выборочных данных модель ошибается максимум на 18кг и 13%;

2) в среднем модель ошибается на 3%;

3) в 80% случаев ошибка не превышала 5%;

4) наибольший разброс ошибок приходится на диапазон 4-6;

5) модель статистически значима и удовлетворяет всем нужным требованиям, кроме гетероскедастичности.

## 3.4 Третье поколение моделей

Путём пошагового отбора переменных были найдены ещё две равнозначные модели ( и ), которые, несмотря на б*о*льшую сложность в сравнении с предыдущими, имеют лучшие оценки при кросс-валидации.

Статистика для модели приведена на рисунках 55-58.

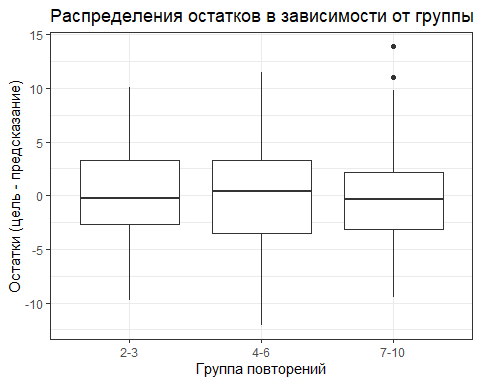


Рисунок 55 – Распределение остатков в зависимости от группы

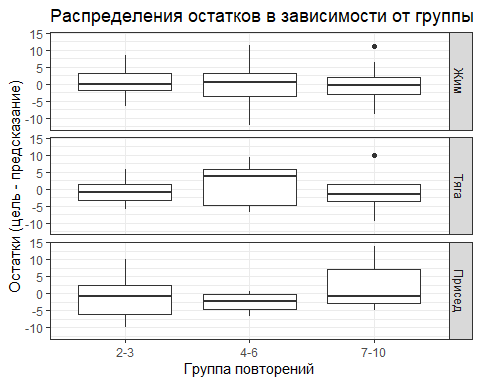


Рисунок 56 – То же самое, разбитое по движению

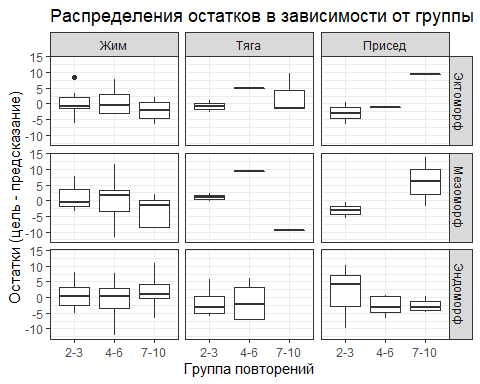


Рисунок 57 – То же самое, разбитое по телосложению

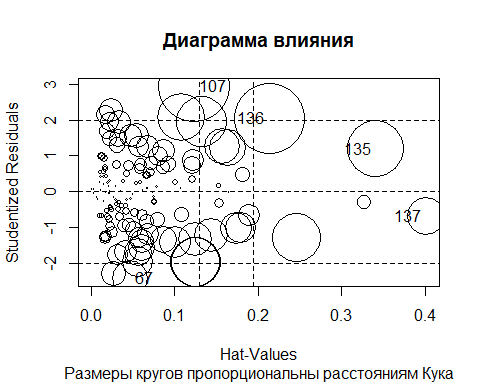


Рисунок 58 – Диаграмма влияния

Основные свойства этой модели:

1) на выборочных данных модель ошибается максимум на 13кг и 12%;

2) в среднем модель ошибается на менее 3%;

3) в 85% случаев ошибка не превышала 5%;

4) в целом, на каждом диапазоне имеется почти одинаковый разброс;

5) модель статистически значима и удовлетворяет всем нужным требованиям.

## 3.5 Сравнение линейных моделей и подведение итогов

В результате перекрёстной проверки для всех моделей при числе блоков 5, 6, 7, 8, 9, 10, 11, 12 было обнаружено, что модель значительно превосходит остальные модели, причём это верно, если делать проверку как на всём диапазоне повторений (2-10), так и на более близком к силовому диапазоне 2-7, см. рисунок 59.

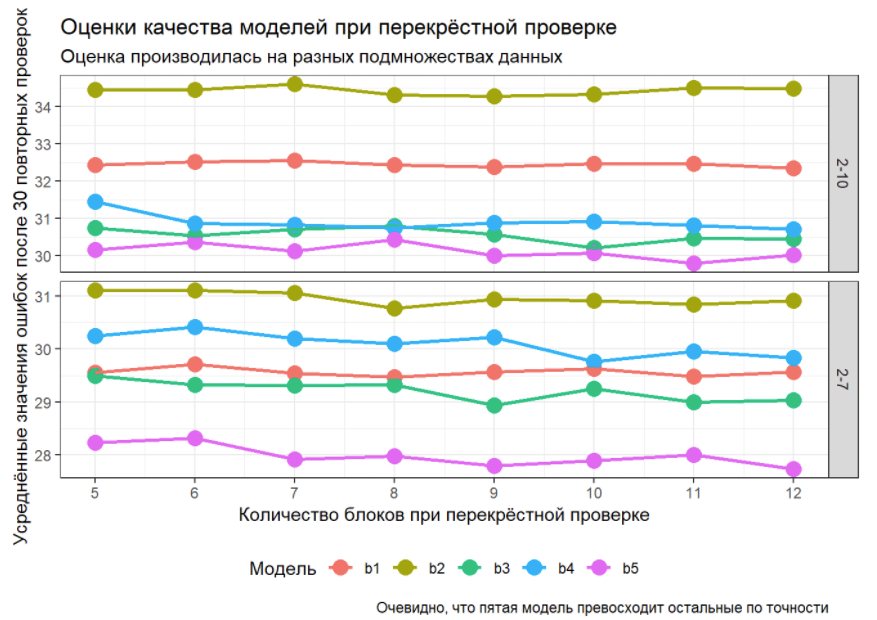


Рисунок 59 – Результаты кросс-валидации по найденным линейным моделям

На втором месте после модели идёт более простая модель .

Что это за модели? Посмотрим на модель . Она имеет вид

где равен 0.9988949 для жима, 1.0129181 для тяги и 1.0362281 для приседа, а есть поправка на диапазон повторений, равная 0.0312223 для диапазона 2-3, 0.0295811 – для 4-6 и 0.0273531 – для 7-10. Как видно, первые три коэффициента близки к единице, однако различия между ними имеют значение; заметим также, что поправка на диапазон повторений 2-3 довольно близка к коэффициенту Вендлера (0.0333).

Кстати, по точности модельпревосходит вторую модель Мориса (ту, что выражена таблицами для каждого движения). Тогда аналогичная таблица, на сколько нужно умножить свой вес для разного числа повторений, выглядит как на рисунке 60.

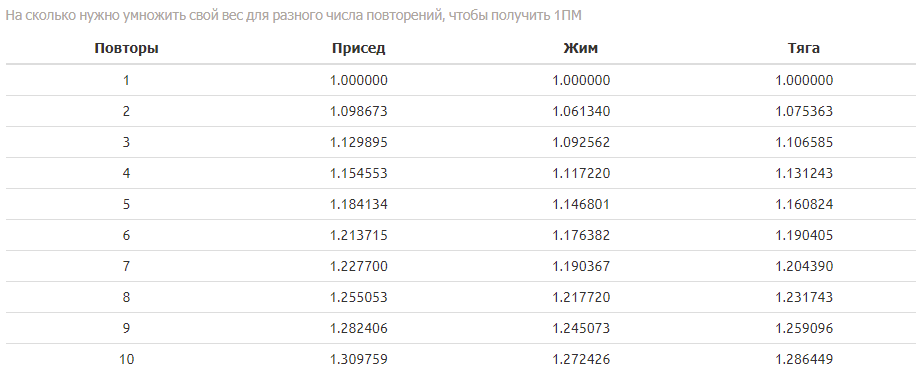


Рисунок 60 – Таблица, на сколько нужно умножать МПМ

Модель имеет два дополнительных слагаемых. Её вид, аналогично:

Здесь при последнем слагаемом равен -0.0000299, – индекс массы тела. Дробь внутри последнего слагаемого можно упростить:

Вообще *существование этого слагаемого (с шестой степенью) кажется чем-то очень сомнительным, но математика показывает, что это имеет смысл*. Результаты дисперсионного анализа (рисунок 61) говорят о том, что включение указанного коэффициента создаёт значимое различие между моделями с ним и без него:

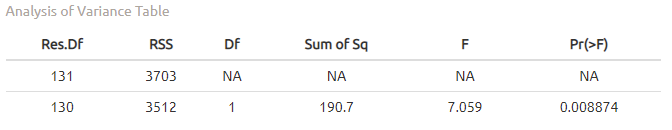


Рисунок 61 – Результаты дисперсионного анализа

Очень интересно, что в этих двух моделях нет зависимости от возраста, типа телосложения, опыта и пола[[26]](#footnote-26). Конечно, нельзя рассмотреть всевозможные вариации таких моделей, но среди сотни рассмотренных (в том числе с помощью пакета *caret*) не было обнаружено доказательств значимости включения этих факторов в модель.

Как итог, для определения ПМ нужны лишь следующие данные:

– МПМ;

– Число повторений;

– Движение;

– Рост и вес спортсмена.

## Нелинейные модели

В этом разделе происходит построение *нелинейных моделей* для предсказания повторных максимумов, проверяются формулы [МакГлотина, Ломбарди](https://ru.wikipedia.org/wiki/Одно_повторение_с_максимальным_весом#МакГлотин_(McGlothin)) и другие нелинейные формулы. Оценка качества моделей происходит по аналогичным принципам, то есть значение имеют:

1) точность моделей при перекрёстной проверке, чтобы их можно было сравнить и выбрать лучшую;

2) точность модели на обучающих данных (низкое смещение), поскольку неточные модели не имеют пользы, даже если превосходят другие при перекрёстной проверке;

3) простота модели.

Сначала рассматриваются отдельные классы моделей из указанной выше ссылки и мои предложения, затем они сравниваются друг с другом и делаются выводы.

### 3.6.1 Основные известные модели

Модели Эпли, Вендлера и О’Коннора – это одна и та же модель вида с разными значениями коэффициента . Как выяснилось, эти отличия обусловлены только тем, на каком разбросе повторений происходил подбор : значение из формулы О’Коннора наиболее оптимально, если формула должна примерно одинаково хорошо работать на не более чем 20 повторениях, а значение из формулы Эпли, предположительно, оптимально для 2-5 повторений (рисунок 62).

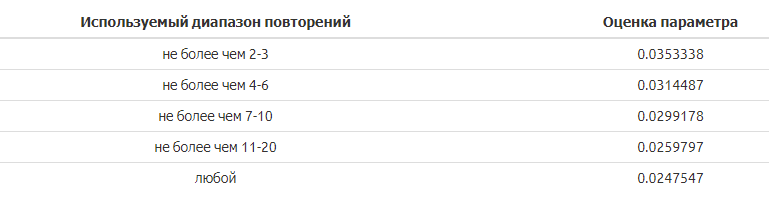


Рисунок 62 – Оценка параметров в модели Эпли

Обучив такую модель (с названием ) для диапазона повторений 2-10, получим результаты в 30% ошибок.

Модели Бжицки и МакГлотина – это тоже одна и та же модель вида

с разными . Обучив такую модель (с названием ) для диапазона повторений 2-10, получим более точную модель (20% ошибок).

Аналогично модели Мэйхью и Ватана — это вариации модели

Обучив такую модель (), получим примерно такие же результаты, как у .

Наконец, модель Ломбарди имеет вид

Обучив её (), получим примерно те же результаты (с параметром, очень близким к 0.1, предложенным Ломбарди):

### 3.6.2 Новые предложения

Кроме этого, я попробую модель :

и :

а также версии предыдущих шести моделей с поправкой коэффициентов на разные факторные переменные и с включением/отсутствием слагаемого (для линейных моделей оно играло роль).

### 3.6.3 Сравнение нелинейных моделей

Сравним названные модели, сделав перекрёстную проверку на диапазоне повторений 2-10 (рисунок 63).

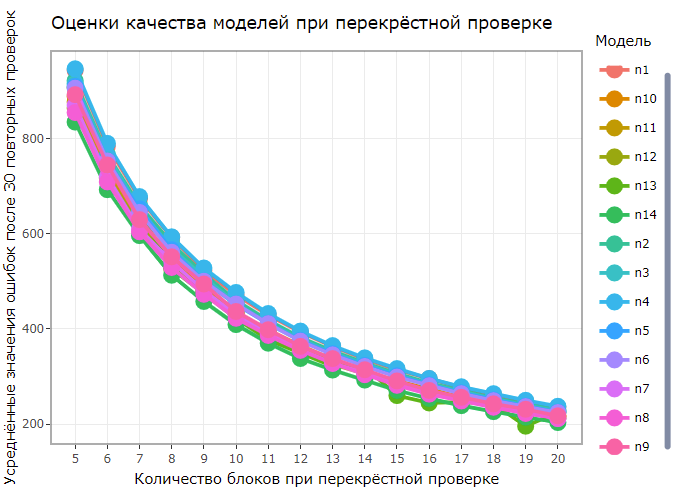


Рисунок 63 – Результаты кросс-валидации по нелинейным моделям

Как видно, самой лучшей моделью оказалась (основанная на моделях Мэйхью и Ватана), на втором месте идёт (моё предложение):

Обе модели имеют одинаковую среднюю ошибку в менее чем 2.9%. По точности они схожи с линейными моделями .

Если сравнить лучшие найденные линейные и нелинейные модели, окажется, что модели и в целом неотличимы, а модель немного превосходит (рисунок 64), но не факт, что значимо.

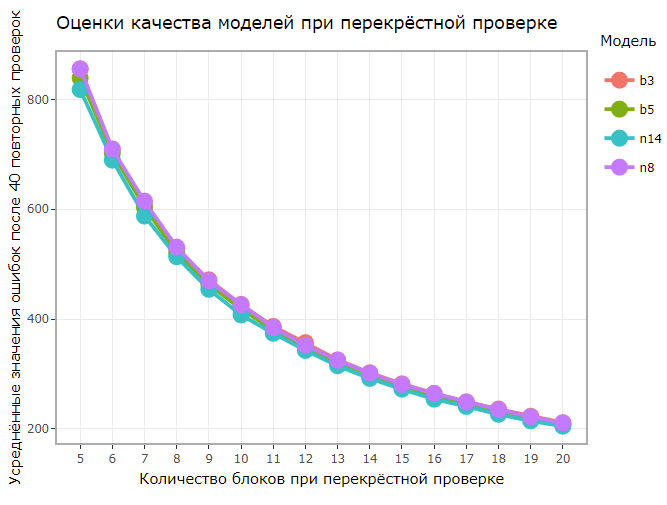


Рисунок 64 – Результаты кросс-валидации по лучшим моделям

# ЗАКЛЮЧЕНИЕ

В результате проведённого исследования была найдена модель, не зависящая от типа телосложения, опыта тренировок и многих других характеристик, но при этом дающая хорошие результаты на 85% данных. Несмотря на относительно большое число слагаемых, эта модель показала одновременно и самое низкое смещение и самую низкую дисперсию, что говорит о высоком качестве.

Я уверен, что большие погрешности в единичных случаях – это та часть закономерности, зависящая от человека и многих неучтённых характеристик. Тем не менее, влияние этих неучтённых факторов далеко не так велико, как предполагали многие: намного больше повторный максимум зависит от обычной математики.

Было продемонстрировано, что для числа повторений выше 10 нельзя построить достаточно точную модель (по крайней мере, при текущем наборе данных), но для диапазона 2-10 — ещё как можно.

Диапазон выше 10 повторений был отсеян по следующим причинам:

1) физиологически менее сильная связь между 1ПМ и большим числом повторений;

2) меньшая детерминированность между 1ПМ и большим числом повторений, обнаруженная при разведочном анализе данных, отчего трудно выявить достоверность данных и построить достаточно точную модель;

3) маленький объём данных, разбросанных очень неравномерно на большом диапазоне (от 11 до 35).

Разумеется, это исследование имеет прикладное назначение. Цель его была не только показать закономерности при решении задачи, но и предоставить само решение, способное помочь реальным людям. С этой целью было создано веб-приложение, с помощью которого можно опробовать полученные модели в реальном времени. Как и во многом другом, создание такого сервиса сопровождалось своими сложностями, однако они уже не имеют прямого отношения к исследованию, как и программная реализация составляющих данной работы, которые были опущены. Посчитать собственные результаты при помощи найденной модели можно по ссылке URL: https://dmitrypasko.shinyapps.io/RMbyMRMestimating/.

Версия этой работы со многими подробностями программной реализации, интерактивными графиками и более расширенной статистикой доступна по ссылке URL: https://rpubs.com/DemetryPascal/RMbyMRMrus\_2.

# 

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Хэтфилд Ф. К. Всестороннее руководство по развитию силы / Красноярск: Союзспорт, 1992. – 284 с.

2 Думай! Или «Супертренинг» без заблуждений (В. Протасенко) [Электронный ресурс] – Режим доступа: <http://athlete.ru/books/protasenko/>, свободный. – Загл. с экрана. (Дата обращения: 02.09.2020).

3 Уикем Х., Гроулмунд Г. Язык R в задачах науки о данных: импорт, обработка, визуализация и моделирование данных.: Пер. с англ. – СПб.: ООО «Диалектика», 2018. – 592 с.

4 Мастицкий С. Э., Шитиков В. К. Статистический анализ и визуализация данных с помощью R. – М.: ДМК Пресс, 2015. – 496 с.

5 Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R. / пер. с англ. Полины А. Волковой. – М: ДМК Пресс, 2016. – 588 с.

6 Джеймс Г., Уиттон Дм Хасти Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С. Э. Мастицкого – М.: ДМК Пресс, 2016. – 450 с:

7 Крупкина, Т. В. Математическая статистика [Электронный ресурс]: курс лекций / Т. В. Крупкина, А. К. Гречкосеев. – Электрон. дан. (3 Мб). – Красноярск: ИПК СФУ, 2009.

8 Brandon Lilly’s Cube Method [Электронный ресурс] – Режим доступа: <https://www.jtsstrength.com/brandon-lillys-cube-method/>, свободный. – Загл. с экрана. (Дата обращения: 21.12.2019).

9 Брантон С. Л., Куц Дж. Н. Анализ данных в науке и технике / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2021. – 574 с.: ил.

10 Калькулятор одноповторного максимума (1ПМ) [Электронный ресурс] – Режим доступа: <https://body1.ru/kalkulyator-odnopovtornogo-maksimuma-1pm/>, свободный. – Загл. с экрана. (Дата обращения: 02.03.2021).

11 Метод Мориса и Райдина или как узнать свой максимум в жиме лежа [Электронный ресурс] – Режим доступа: <https://power-fitness.ru/metod-morisa-i-rajdina-ili-kak-uznat-svoj-maksimum-v-zhime-lezha.html>, свободный. – Загл. с экрана. (Дата обращения: 08.03.2021).

12 Расчёт разового максимума в жиме лёжа [Электронный ресурс] – Режим доступа: <http://frs24.ru/st/maksimum-v-zhime-lezha-raschet/>, свободный. – Загл. с экрана. (Дата обращения: 02.01.2021).

13 Одно повторение с максимальным весом [Электронный ресурс] – Режим доступа: <https://ru.wikipedia.org/wiki/Одно_повторение_с_максимальным_весом>, свободный. – Загл. с экрана. (Дата обращения: 02.24.2021).

14 National Strength and Conditioning Association (NSCA) [Электронный ресурс] – Режим доступа: <https://www.nsca.com>, свободный. – Загл. с экрана. (Дата обращения: 02.04.2021).

1. иногда их называют *разовыми максимумами*, *РМ* [↑](#footnote-ref-1)
2. имеется в виду, что нетрудно и более безопасно на следующей тренировке взять достаточно большой вес и сделать с ними максимальное число повторений (около 4-8) [↑](#footnote-ref-2)
3. зачастую так думают, люди, которые своему опыту доверяют намного больше, чем научному знанию, тем более в спортивном мире часто всплывают разные псевдонаучные методики, которые приносят спортсменам только вред [↑](#footnote-ref-3)
4. URL: <https://docs.google.com/forms/d/e/1FAIpQLSeownLfmO9Sx3rwI9gA5Msm0Z_xiFbMj12QHS490tm74UfFJw/viewform> [↑](#footnote-ref-4)
5. URL: <https://docs.google.com/forms/d/e/1FAIpQLSfDJEK3SxuQnVlD-_mHbFvB47BpUouUmUv_md10gSQcnsI3TA/viewform> [↑](#footnote-ref-5)
6. «преимущественно» означает, что предположительно именно из этих группы поступила основная масса опрошенных [↑](#footnote-ref-6)
7. URL: <https://vk.com/powerliftingnews> [↑](#footnote-ref-7)
8. URL: <https://vk.com/ilovepowerlifting> [↑](#footnote-ref-8)
9. URL: <https://vk.com/powerliftingworld> [↑](#footnote-ref-9)
10. URL: <https://vk.com/tvoytrenercom> [↑](#footnote-ref-10)
11. URL: <https://www.facebook.com/PowerliftingMotivation/> [↑](#footnote-ref-11)
12. URL: <https://www.facebook.com/groups/IPLPowerlifting/> [↑](#footnote-ref-12)
13. URL: <https://www.facebook.com/groups/powerliftingmotivation/> [↑](#footnote-ref-13)
14. URL: <https://www.facebook.com/groups/powerliftingmemes/> [↑](#footnote-ref-14)
15. URL: <https://www.rstudio.com> [↑](#footnote-ref-15)
16. если будет доказано наличие гиперплазии (деления мышечных волокон) в скелетных мышцах человека, можно будет утверждать, что силовые тренировки увеличивают количество белых волокон в работающих мышцах [↑](#footnote-ref-16)
17. речь идёт о тренировках с более чем 90% от ПМ или с более чем 100% от ПМ в частичной амплитуде, а также об изометрических упражнениях; во многом такие упражнения выполняются для того, чтобы приучить организм поднимать огромные веса, что и значит – тренировка способности генерировать б**о**льший импульс и увеличение порога его подавления [↑](#footnote-ref-17)
18. в некотором смысле это одно и то же [↑](#footnote-ref-18)
19. конечно, это лишь условные размышления без учёта особенностей отдельных частей тела и того, что ввиду усталости каждое следующее повторение делается дольше предыдущего. Важно лишь усвоить, что разные диапазоны повторений по-разному связаны с силой. [↑](#footnote-ref-19)
20. дело в том, что одни и те же люди могли вносить более одного ответа. В итоге два наблюдения считаются принадлежащими одному человеку, если они совпадают по переменным *AgeGroup*, *Height*, *BodyType*, *Experience*, *Sex*, *IndexGroup* (которые вряд ли изменятся за время между фиксацией разных наблюдений) [↑](#footnote-ref-20)
21. URL: <https://ru.wikipedia.org/wiki/Индекс_массы_тела> [↑](#footnote-ref-21)
22. к сожалению, этого не получилось сделать, так как испытуемые редко пользуются электронной почтой, видимо [↑](#footnote-ref-22)
23. URL на исходный файл с данными <https://github.com/PasaOpasen/Powerlifting-training-diary-and-articles/blob/master/Estimating%20RM/data.tsv> [↑](#footnote-ref-23)
24. например, высокий импульс ЦНС, не соответствующий уровню креатинфосфата. В таком случае человек может на раз пожать большой вес, но на несколько раз уже *недостаточно большой*: при ПМ в 140 на 3 раза жмётся не почти 130, а 120 [↑](#footnote-ref-24)
25. такую нижнюю границу можно предсказать, просто умножив вес примерно на 1.5, исследование же рассчитано на поиск верхней границы [↑](#footnote-ref-25)
26. по крайней мере, наблюдения от женщин, прошедших опрос, не становились выбросами для этих моделей, а хорошо вписывались во многие наблюдения, принадлежавшие мужчинам [↑](#footnote-ref-26)