

# Отчет о проверке на заимствования №1



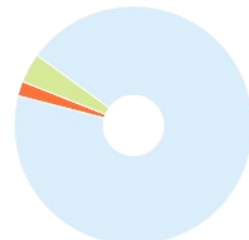
**Автор:** Пасько Д. А.  
**Проверяющий:** user 3 4  
**Организация:** Кубанский Государственный университет  
Отчет предоставлен сервисом «Антиплагиат» - <http://kubsu.antiplagiat.ru>

## ИНФОРМАЦИЯ О ДОКУМЕНТЕ

№ документа: 267  
Начало загрузки: 03.06.2021 12:44:24  
Длительность загрузки: 00:00:13  
Имя исходного файла: ВКР - 2021 - 02.04.01 - Пасько Д. А..pdf  
Название документа: ВКР - 2021 - 02.04.01 - Пасько Д. А.  
Размер текста: 1 кБ  
Тип документа: Магистерская диссертация  
Символов в тексте: 50744  
Слов в тексте: 6409  
Число предложений: 296

## ИНФОРМАЦИЯ ОБ ОТЧЕТЕ

Последний готовый отчет (ред.)  
Начало проверки: 03.06.2021 12:44:37  
Длительность проверки: 00:00:42  
Комментарии: не указано  
Поиск с учетом редактирования: да  
Модули поиска: ИПС Адилет, Библиография, Сводная коллекция ЭБС, Интернет Плюс, Сводная коллекция РГБ, Цитирование, Переводные заимствования (RuEn), Переводные заимствования по eLIBRARY.RU (EnRu), Переводные заимствования по Интернету (EnRu), Переводные заимствования издательства Wiley (RuEn), eLIBRARY.RU, СПС ГАРАНТ, Модуль поиска "КубГУ", Медицина, Диссертации НББ, Перефразирования по eLIBRARY.RU, Перефразирования по Интернету, Патенты СССР, РФ, СНГ, Шаблонные фразы, Кольцо вузов, Издательство Wiley, Переводные заимствования



### ЗАИМСТВОВАНИЯ

1,58% ■

### САМОЦИТИРОВАНИЯ

0% ■

### ЦИТИРОВАНИЯ

4,1% ■

### ОРИГИНАЛЬНОСТЬ

94,32% ■

Заимствования — доля всех найденных текстовых пересечений, за исключением тех, которые система отнесла к цитированиям, по отношению к общему объему документа.  
Самоцитирования — доля фрагментов текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника, автором или соавтором которого является автор проверяемого документа, по отношению к общему объему документа.  
Цитирования — доля текстовых пересечений, которые не являются авторскими, но система посчитала их использование корректным, по отношению к общему объему документа. Сюда относятся оформленные по ГОСТу цитаты; общеупотребительные выражения; фрагменты текста, найденные в источниках из коллекций нормативно-правовой документации.  
Текстовое пересечение — фрагмент текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника.  
Источник — документ, проиндексированный в системе и содержащийся в модуле поиска, по которому проводится проверка.  
Оригинальность — доля фрагментов текста проверяемого документа, не обнаруженных ни в одном источнике, по которым шла проверка, по отношению к общему объему документа.  
Заимствования, самоцитирования, цитирования и оригинальность являются отдельными показателями и в сумме дают 100%, что соответствует всему тексту проверяемого документа.  
Обращаем Ваше внимание, что система находит текстовые пересечения проверяемого документа с проиндексированными в системе текстовыми источниками. При этом система является вспомогательным инструментом, определение корректности и правомерности заимствований или цитирований, а также авторства текстовых фрагментов проверяемого документа остается в компетенции проверяющего.

№	Доля в отчете	Доля в тексте	Источник	Актуален на	Модуль поиска	Блоков в отчете	Блоков в тексте	Комментарии
[01]	2,73%	2,73%	не указано	раньше 2011	Библиография	1	1	
[02]	1,09%	1,35%	не указано	раньше 2011	Шаблонные фразы	12	14	
[03]	0,59%	1,07%	Гомлешхова М.К. Подбор,расстановка и оценка эффективности работы кадров банков	05 Ноя 2019	Модуль поиска "КубГУ"	2	2	
[04]	0%	0,73%	Д Ханагуа С.З. Маркетинговая стратегия для предприятий табачной промышленности	03 Июн 2019	Модуль поиска "КубГУ"	0	2	
[05]	0%	0,7%	p_2132_178.pdf	04 Апр 2017	Кольцо вузов	0	2	
[06]	0%	0,69%	ВКР_КарасьНВ_17ФИИТ(м)ИС_250619.docx	25 Июн 2019	Кольцо вузов	0	2	
[07]	0%	0,69%	<a href="https://www.kubsau.ru/upload/iblock/cfe/cfe788d66932cb02cfd2267ca0cf5519.pdf">https://www.kubsau.ru/upload/iblock/cfe/cfe788d66932cb02cfd2267ca0cf5519.pdf</a> <a href="https://kubsau.ru">https://kubsau.ru</a>	14 Сен 2018	Интернет Плюс	0	3	
[08]	0,14%	0,68%	производ. практика <a href="http://csu.ru">http://csu.ru</a>	29 Янв 2017	Перефразирования по Интернету	1	2	
[09]	0%	0,66%	Лебедева Ю.Ю. Моделирование систем управления организации	15 Июн 2020	Модуль поиска "КубГУ"	0	1	
[10]	0%	0,52%	Исследование качества вращения привода в навигационной системе <a href="http://elib2.altstu.ru">http://elib2.altstu.ru</a>	08 Сен 2017	Интернет Плюс	0	3	
[11]	0,5%	0,5%	<a href="http://metior.ru/upload/002/u201/000/menedzhment-kulikova-zajarnaja.pdf">http://metior.ru/upload/002/u201/000/menedzhment-kulikova-zajarnaja.pdf</a> <a href="http://metior.ru">http://metior.ru</a>	27 Окт 2020	Интернет Плюс	3	3	
[12]	0%	0,49%	suchkova_k_s_statisticheskoe-issledovanie-sostoyaniya-i-razvitiya-jilishchno-kommunalnogo-hozyaystva-v-rossii-i-e.docx	22 Мар 2019	Кольцо вузов	0	2	
[13]	0%	0,49%	ПРИМЕНЕНИЕ МЕТОДОВ КЛАССИФИКАЦИИ И КЛАСТЕРИЗАЦИИ ДЛЯ АНАЛИЗА ДАННЫХ С ВОДНЫХ	05 Июн 2019	Кольцо вузов	0	1	

			ОБЪЕКТОВ				
[14]	0%	0,49%	ПРИМЕНЕНИЕ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ ДЛЯ ПРОГНОЗА ВЕРОЯТНОСТИ ПРИНАДЛЕЖНОСТИ ОБЪЕКТА К ТОМУ ИЛИ ИНОМУ КЛАССУ ЗАГРЯЗНЕНИЯ ВОДЫ	05 Июн 2019	Кольцо вузов	0	1
[15]	0%	0,45%	SVD-скатие, модификации <a href="https://knowledge.allbest.ru">https://knowledge.allbest.ru</a>	30 Мая 2021	Интернет Плюс	0	3
[16]	0%	0,44%	Статистические пакеты программ в социально-экономических исследованиях – тема научной статьи по компьютерным и информационным наукам читайте бесплатно текст научно-исследовательской работы в электронной библиотеке КиберЛенинка <a href="https://cyberleninka.ru">https://cyberleninka.ru</a>	12 Окт 2020	Интернет Плюс	0	3
[17]	0%	0,42%	Таков Р.М. Развитие рынка венчурных инвестиций как фактор повышения экономической безопасности РФ	10 Июн 2019	Модуль поиска "КубГУ"	0	1
[18]	0%	0,42%	Д Санакоев А.В. Подходы к стандартизации государственных и муниципальных услуг	14 Июн 2019	Модуль поиска "КубГУ"	0	1
[19]	0%	0,42%	Д Санакоев А.В. Подходы к стандартизации государственных и муниципальных услуг	16 Июн 2019	Модуль поиска "КубГУ"	0	1
[20]	0%	0,42%	Фурсов Е.А. Обеспечение экономической безопасности внешнеэкономической деятельности региона	20 Ноя 2019	Модуль поиска "КубГУ"	0	1
[21]	0%	0,42%	Д Кобяков В.А. Анализ удовлетворенности потребителей как инструмент управления качеством	23 Июн 2020	Модуль поиска "КубГУ"	0	1
[22]	0%	0,4%	Клименко, Павел Валерьевич Интеграция традиций и инноваций как ресурс патриотического воспитания курсантов военного вуза : диссертация ... кандидата педагогических наук : 13.00.01 Краснодар 2019 <a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	19 Авг 2020	Сводная коллекция РГБ	0	1
[23]	0%	0,36%	ПРОГРАММА ПРАКТИКИ НАПРАВЛЕНИЕ ПОДГОТОВКИ / СПЕЦИАЛЬНОСТЬ (код и наименование направления подготовки) - PDF Free Download <a href="https://docplayer.ru">https://docplayer.ru</a>	01 Окт 2020	Интернет Плюс	0	2
[24]	0%	0,34%	$\Delta T = \Delta P = P \cdot K \cdot P \cdot c, (2)$ - PDF Free Download <a href="https://docplayer.ru">https://docplayer.ru</a>	16 Окт 2020	Интернет Плюс	0	1
[25]	0%	0,32%	"Арабская весна" в Египте: причины и последствия: выпускная квалификационная работа бакалавра <a href="http://biblioclub.ru">http://biblioclub.ru</a>	21 Янв 2020	Сводная коллекция ЭБС	0	1
[26]	0%	0,3%	Ààîë òàñîðàüò çàààîëëë à ñëñòàâ UniTest <a href="http://topreferat.znate.ru">http://topreferat.znate.ru</a>	29 Янв 2017	ПЕРЕФРАЗИРОВАНИЯ ПО ИНТЕРНЕТУ	0	1
[27]	0%	0,29%	Г. Джеймс, Д. Уиттон, Т. Хастинс, Р. Тибишани; Пер. с англ. С. Э. Мастицкого Введение в статистическое обучение с примерами на языке R Москва 2016 <a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	05 Авг 2019	Сводная коллекция РГБ	0	1
[28]	0,28%	0,28%	Об утверждении состава рабочей группы по выработке предложений по формированию цифрового пространства Евразийского экономического союза - ИПС "Эдilet" <a href="http://adilet.zan.kz">http://adilet.zan.kz</a>	04 Окт 2017	ИПС Адилет	1	1
[29]	0%	0,28%	Туристско-рекреационный потенциал как основа развития въездного туризма Китая <a href="http://dep.nlb.by">http://dep.nlb.by</a>	16 Янв 2020	Диссертации НББ	0	1
[30]	0,08%	0,27%	Проект участка производства электролитического медного порошка производительностью 100 тонн медного порошка в год <a href="https://nauchkor.ru">https://nauchkor.ru</a>	29 Окт 2020	Интернет Плюс	1	2
[31]	0%	0,26%	Магистерская диссертация <a href="http://studentlibrary.ru">http://studentlibrary.ru</a>	19 Дек 2016	Медицина	0	1
[32]	0%	0,25%	R в действии. Анализ и визуализация данных в программе R <a href="https://e.lanbook.com">https://e.lanbook.com</a>	21 Янв 2020	Сводная коллекция ЭБС	0	1
[33]	0%	0,25%	58703 <a href="http://e.lanbook.com">http://e.lanbook.com</a>	09 Мар 2016	Сводная коллекция ЭБС	0	1
[34]	0%	0,25%	hirina_d_v_pochemu-izbiratel-vybiraet-alternativu-dlya-germanii-radikalnye-izmeneniya-predpochteniy-izbirateley.docx	12 Мая 2019	Кольцо вузов	0	1

[35]	0%	0,25%	Детярев В.П.	09 Июн 2018	Кольцо вузов	0	1	
[36]	0,04%	0,25%	Биотехносфера № 05-06.2012 <a href="http://studentlibrary.ru">http://studentlibrary.ru</a>	19 Дек 2016	Медицина	1	1	
[37]	0%	0,24%	Проекционные алгоритмы вихревых 2D течений в сложных областях. <a href="http://elibrary.ru">http://elibrary.ru</a>	04 Авг 2016	eLIBRARY.RU	0	2	
[38]	0%	0,23%	Потапенко, Александр Сергеевич Совершенствование тепловых процессов в установке непрерывного совмещенного литья и прессования цветных металлов : диссертация ... кандидата технических наук : 05.14.04 Красноярск 2018 <a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	22 Фев 2019	Сводная коллекция РГБ	0	1	
[39]	0,23%	0,23%	Strength Training - Kinesiology 241 with Herman at University of Michigan - Ann Arbor - StudyBlue <a href="https://studyblue.com">https://studyblue.com</a>	09 Янв 2018	Переводные заимствования (RuEn)	1	1	
[40]	0%	0,23%	Прогнозирование финансово-экономических показателей по разнородным данным <a href="https://book.ru">https://book.ru</a>	21 Янв 2020	Сводная коллекция ЭБС	0	1	
[41]	0%	0,23%	Статистический анализ и визуализация данных с помощью R <a href="https://e.lanbook.com">https://e.lanbook.com</a>	22 Янв 2020	Сводная коллекция ЭБС	0	1	
[42]	0%	0,23%	Пушкина, Татьяна Анатольевна Повышение физической работоспособности высококвалифицированных спортсменов на основе применения комплексного антиоксидантного продукта : диссертация ... кандидата биологических наук : 03.03.01 Москва 2020 <a href="http://dlib.rsl.ru">http://dlib.rsl.ru</a>	22 Окт 2020	Сводная коллекция РГБ	0	1	
[43]	0%	0,22%	Тканеинженерные конструкции на основе синтетического каркаса и недифференцированных костномозговых мононуклеарных клеток в лечении стенозов трахеи <a href="http://ksma.ru">http://ksma.ru</a>	03 Фев 2021	Интернет Плюс	0	1	
[44]	0%	0,21%	С. Г.Тарайцева, Т. В.Юрченко РАЗРАБОТКА ПРИКЛАДНЫХ РЕШЕНИЙ НА ПЛАТФОРМЕ 1С: ПРЕДПРИЯТИЕ 8. Учебное пособие - PDF Free Download <a href="https://docplayer.ru">https://docplayer.ru</a>	15 Дек 2020	Интернет Плюс	0	1	
[45]	0%	0,2%	<a href="http://theses2.sportedu.ru/sites/theses2.sportedu.ru/files/dissertaciya_pushkina.pdf">http://theses2.sportedu.ru/sites/theses2.sportedu.ru/files/dissertaciya_pushkina.pdf</a> <a href="http://theses2.sportedu.ru">http://theses2.sportedu.ru</a>	18 Окт 2020	Интернет Плюс	0	2	
[46]	0%	0,2%	Построение эконометрических моделей и анализ факторов экспорта и импорта Санкт-Петербурга   Статья в журнале «Молодой ученый» <a href="https://moluch.ru">https://moluch.ru</a>	07 Мар 2021	Интернет Плюс	0	2	
[47]	0%	0,2%	ГЛАВА 1 Введение   Анализ временных рядов с помощью R <a href="https://ranalytics.github.io">https://ranalytics.github.io</a>	21 Дек 2020	Интернет Плюс	0	1	
[48]	0%	0,2%	Программные средства научных исследований. <a href="http://elibrary.ru">http://elibrary.ru</a>	14 Окт 2019	Перефразирования по eLIBRARY.RU	0	1	
[49]	0%	0,2%	struk_a_v_regionalnye-razlichiya-v-identichnosti-i-politicheskikh-vzglyadah-ukrainskogo-naseleniya.docx	10 Окт 2016	Кольцо вузов	0	1	
[50]	0%	0,2%	hohoev_a_p_risk-menedjment-v-torgovyh-strategiyah.docx	04 Июн 2017	Кольцо вузов	0	1	
[51]	0%	0,2%	hohoev_a_p_risk-menedjment-torgovyh-strategy-v-usloviyah-neeftivnogo-rynka.pdf	20 Мая 2018	Кольцо вузов	0	1	
[52]	0%	0,19%	ГОСТ 21.101-97 Система проектной документации для строительства (СПДС). Основные требования к проектной и рабочей документации, ГОСТ от 29 декабря 1997 года №21.101-97, <a href="http://docs.cntd.ru">http://docs.cntd.ru</a>	03 Июн 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[53]	0%	0,19%	ГОСТ 21.101-97 Система проектной документации для строительства (СПДС). Основные требования к проектной и рабочей документации, ГОСТ от 29 декабря 1997 года №21.101-97, <a href="http://docs.cntd.ru">http://docs.cntd.ru</a>	03 Июн 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[54]	0%	0,19%	ДИФфуЗИЯ ВЕЩЕСТВ <a href="https://school-science.ru">https://school-science.ru</a>	28 Апр 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.

[55]	<div><div></div></div> 0%	0,19%	ГОСТ 21.101-97 Система проектной документации для строительства (СПДС). Основные требования к проектной и рабочей документации, ГОСТ от 29 декабря 1997 года №21.101-97, <a href="http://docs.cntd.ru/243_21_61_0_0.600_63767033">http://docs.cntd.ru/243_21_61_0_0.600_63767033</a>	03 Июн 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[56]	<div><div></div></div> 0%	0,17%	Социология: Курс лекций для студентов всех направлений и специальностей / Под ред. В.А. Михайлова. - Ульяновск: УлГТУ, 2004. - 207 с. <a href="http://venec.ulstu.ru">http://venec.ulstu.ru</a>	16 Апр 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[57]	<div><div></div></div> 0%	0,16%	274604 <a href="http://biblioclub.ru">http://biblioclub.ru</a>	раньше 2011	Сводная коллекция ЭБС	0	1	Источник исключен. Причина: Маленький процент пересечения.
[58]	<div><div></div></div> 0%	0,16%	Физическая культура, спорт и здоровье студенческой молодежи в современных условиях (1/2) <a href="http://rsvpu.ru">http://rsvpu.ru</a>	04 Ноя 2018	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[59]	<div><div></div></div> 0%	0,16%	МЕТОДИКА РАЗВИТИЯ СИЛОВЫХ СПОСОБНОСТЕЙ У ЮНОШЕЙ 15-16 ЛЕТ НА НАЧАЛЬНОМ ЭТАПЕ ОБУЧЕНИЯ В ПАУЭРЛИФТИНГЕ - Студенческий научный форум <a href="https://scienceforum.ru">https://scienceforum.ru</a>	09 Фев 2019	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[60]	<div><div></div></div> 0%	0,13%	Дипломная работа: Из опыта подготовки и выполнения норматива мастера спорта России по пауэрлифтингу за период обучения в вузе - BestReferat.ru <a href="https://bestreferat.ru">https://bestreferat.ru</a>	27 Окт 2020	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[61]	<div><div></div></div> 0%	0,12%	К ПРИМЕНЕНИЮ ГОСТ 7.32-2017 В ВУЗЕ. ТРЕБОВАНИЯ. <a href="http://elibrary.ru">http://elibrary.ru</a>	10 Фев 2020	eLIBRARY.RU	0	1	Источник исключен. Причина: Маленький процент пересечения.
[62]	<div><div></div></div> 0%	0,12%	ЮРИДИЧЕСКАЯ ДЕФЕКТОЛОГИЯ В СИСТЕМЕ ГУМАНИТАРНОГО ЗНАНИЯ. ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА О ДИССЕРТАЦИИ КОЖОКАРЯ ИГОРЯ ПЕТРОВИЧА НА ТЕМУ «ТЕХНИКО-ЮРИДИЧЕСКИЕ ДЕФЕКТЫ В РОССИЙСКОМ ПРАВЕ» (М., 2020. 417 С.), ПРЕДСТАВЛЕННОЙ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ ДОКТОРА ... <a href="http://elibrary.ru">http://elibrary.ru</a>	01 Фев 2021	eLIBRARY.RU	0	1	Источник исключен. Причина: Маленький процент пересечения.
[63]	<div><div></div></div> 0%	0,06%	не указано	раньше 2011	Цитирование	0	1	Источник исключен. Причина: Маленький процент пересечения.

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
1 Актуальность исследования .....	5
1.1 Зачем предсказывать повторные максимумы? .....	5
1.2 О калькуляторах .....	7
1.3 О данных и ресурсах .....	8
1.4 Отправная точка .....	10
1.5 Гипотеза и допущения .....	11
1.6 Соображения о возможности построения модели .....	15
1.7 План исследования .....	17
2 Описание выборки и разведочный анализ данных .....	18
2.1 О прошедших опрос .....	18
2.2 Взаимодействия переменных .....	26
2.3 Исправление опечаток .....	37
2.4 Удалённые данные .....	38
2.5 Подведение итогов и отбор признаков .....	43
3 Построение моделей .....	46
3.1 Недостатки исходного решения .....	46
3.2 Уточнение коэффициентов .....	49
3.3 Уточнение коэффициентов: зависимость от диапазона .....	52
3.4 Третье поколение моделей .....	56
3.5 Сравнение линейных моделей и подведение итогов .....	59
3.6 Нелинейные модели .....	63
3.6.1 Основные известные модели .....	63
3.6.2 Новые предложения .....	65
3.6.3 Сравнение нелинейных моделей .....	65
ЗАКЛЮЧЕНИЕ .....	68
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	70

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<sup>1</sup>*), основываясь на результатах, которые ты показывал совсем недавно в процессе обычных тренировок или которые намного проще продемонстрировать<sup>2</sup> (*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:

---

<sup>1</sup> иногда их называют *разовыми максимумами, РМ*

<sup>2</sup> 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)

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. статью по URL <https://body1.ru/kalkulyator-odnopolvornogo-maksimuma-1pm/>);

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: пытаться 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<sup>3</sup>; 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 (<https://power-fitness.ru/metod-morisa-i-rajdina-ili-kak-uznat-svoj-maksimum-v-zhime-lezha.html>), которые, как позже будет показано, нерабочие и, скорее всего, получены из слишком маленьких выборок;

---

<sup>3</sup> 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

4) формулы О Коннора, Бжицки (Бржыки), Лэндера (собранные вместе по <http://frs24.ru/st/maksimum-v-zhime-lezha-raschet/>) и другие, с которыми можно ознакомиться в Википедии по ссылке [https://ru.wikipedia.org/wiki/Одно\\_повторение\\_с\\_максимальным\\_весом;](https://ru.wikipedia.org/wiki/Одно_повторение_с_максимальным_весом;)

5) также калькуляторы, представленные на сайтах [https://powerliftingrating.ru/repeat\\_calculator](https://powerliftingrating.ru/repeat_calculator), <http://evgeniyfit.ru/fitness-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<sup>4</sup> на русском языке и его англоязычный аналог<sup>5</sup>.

---

<sup>4</sup>

[https://docs.google.com/forms/d/e/1FAIpQLSeownLfmO9Sx3rwI9gA5Msm0Z\\_xiFbMj12QHS490tm74UfFJw/viewform](https://docs.google.com/forms/d/e/1FAIpQLSeownLfmO9Sx3rwI9gA5Msm0Z_xiFbMj12QHS490tm74UfFJw/viewform)

<sup>5</sup>

[https://docs.google.com/forms/d/e/1FAIpQLSfDJEK3SxuQnVID-mHbFvB47BpUouUmUv\\_md10gSQcnsI3TA/viewform](https://docs.google.com/forms/d/e/1FAIpQLSfDJEK3SxuQnVID-mHbFvB47BpUouUmUv_md10gSQcnsI3TA/viewform)

Русскоязычный опрос распространялся через соц. сеть ВКонтакте преимущественно<sup>6</sup> в следующих группах: «Пауэрлифтинг | Тяжелая атлетика»<sup>7</sup>, «Я ♥ ПАУЭРЛИФТИНГ»<sup>8</sup>, «ПАУЭРЛИФТИНГ И ЖИМ ЛЁЖА 18+»<sup>9</sup>, «Твой Тренер»<sup>10</sup>; аналогично англоязычный опрос распространялся на Facebook преимущественно здесь: «Powerlifting Motivation»<sup>11</sup>, «International Powerlifting League (IPL)»<sup>12</sup>, «Powerlifting Motivation Chat»<sup>13</sup>, «Powerlifting Memes»<sup>14</sup>. 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<sup>15</sup> 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 книгами [3,4,5,6].

---

<sup>6</sup> “преимущественно” означает, что предположительно именно из этих группы поступила основная масса опрошенных

<sup>7</sup> <https://vk.com/powerliftingnews>

<sup>8</sup> <https://vk.com/ilovepowerlifting>

<sup>9</sup> <https://vk.com/powerliftingworld>

<sup>10</sup> <https://vk.com/tvoytrenecom>

<sup>11</sup> <https://www.facebook.com/PowerliftingMotivation/>

<sup>12</sup> <https://www.facebook.com/groups/IPLPowerlifting/>

<sup>13</sup> <https://www.facebook.com/groups/powerliftingmotivation/>

<sup>14</sup> <https://www.facebook.com/groups/powerliftingmemes/>

<sup>15</sup> <https://www.rstudio.com>

## 1.4 Отправная точка

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

предполагаемый повторный максимум = вес \* (1 + 0.0333 \* повторения), (1)

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

$$RM = WEIGHT \cdot (1 + 0.0333 \cdot REPS). \quad (2)$$

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.

### **1.5 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 большую 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 большую 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<sup>16</sup>.

4) Количество вовлечённых в работу волокон. Чем их больше, тем больше сила мышц. Количество задействованных волокон зависит от

---

<sup>16</sup> 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

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<sup>17</sup>.

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 <sup>2</sup>одноповторный максимум можно с небольшими ошибками предсказать через многоповторный при помощи модели вида

---

<sup>17</sup> 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 больший импульс и увеличение порога его подавления



$$RM = MRM \cdot (x + y \cdot REPS) + f(MRM, HEIGHT, WEIGHT) + \varepsilon \quad (3)$$

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

Здесь (формула (3))  $x, y$  – некоторые числа,  $f$  – функция, которую ещё придётся подобрать,  $\varepsilon$  – 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  $50 * 6 \approx 60$  (вес 50 на 6 раз примерно значит 60 на раз) должно следовать  $100 * 6 \approx 120$  с, возможно, небольшими поправками (до 10%) на другие факторы (функция  $f$ ). 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<sup>18</sup> (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 рН).

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 Вадима Протасенко [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, алактатный 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<sup>19</sup>.

---

<sup>18</sup> в некотором смысле это одно и то же

<sup>19</sup> конечно, это лишь условные размышления без учёта особенностей отдельных частей тела и того, что ввиду усталости каждое следующее повторение делается дольше

## 1.6 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

---

предыдущего. Важно лишь усвоить, что разные диапазоны повторений по-разному связаны с силой.

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.**<sup>2</sup> 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 настроения; при следующей тренировке эти параметры могут быть уже значимо другими, отчего правила, хорошо работавшие недавно, перестанут хорошо работать сейчас.

**В то же время имеются**<sup>2</sup> 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<sup>20</sup>; они содержатся в файле по ссылке [https://github.com/PasaOpasen/Powerlifting-training-diary-and-articles/blob/master/Estimating%20RM/data\(rus\).tsv](https://github.com/PasaOpasen/Powerlifting-training-diary-and-articles/blob/master/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

---

<sup>20</sup> 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)

- *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, *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 Рис. 1.

	mean	sd	median	mad	min	max	range
RM	154.121	51.089	147.500	48.184	20.000	300.000	280.000
MRM	127.158	42.675	120.000	40.771	15.000	250.000	235.000
Count	8.144	7.071	6.000	4.448	2.000	35.000	33.000
Age	28.017	8.990	27.000	8.896	15.000	62.000	47.000
Weight	89.608	18.883	87.000	17.050	47.000	160.000	113.000
Height	176.931	8.252	177.000	8.896	155.000	198.000	43.000
Index	28.474	5.011	27.767	4.708	18.827	45.269	26.442

Action <fct>	Sex <fct>	Experience <fct>	BodyType <fct>	CountGroup <fct>	AgeGroup <fct>
X Жим :124	Мужчина:165	До двух лет :31	Эктоморф:36	2-3 :47	<20 :31
X.1 Тяга : 26	Женщина: 9	2-3 года :31	Мезоморф:52	4-6 :51	20-27:60
X.2 Присед: 24	NA	4-5 лет :47	Эндоморф:86	7-10 :41	28-35:54
X.3 NA	NA	6-10 лет :37	NA	11-20:23	>35 :29
X.4 NA	NA	больше 10 лет:28	NA	>20 :12	NA

Рисунок 1

Сделаем некоторые выводы о выборке. Почти все испытуемые – мужчины (Рис. 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

Среди испытуемый почти половину составляли эндоморфы (Рис. 3)

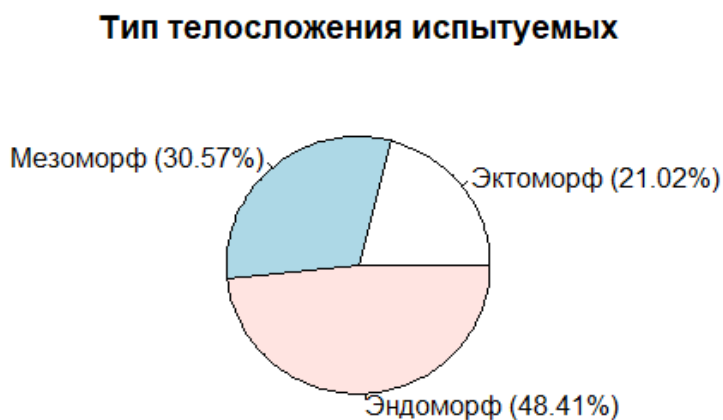


Рисунок 3



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

Chi-squared test for given probabilities:

```
data.unique %>% filter(Sex == "Мужчина") %>% select(BodyType) %>% table()
```

Test statistic	df	P value
16.28	2	0.0002914 ***

Рисунок 4

Большее половины наблюдений относятся к жиму лёжа (Рис. 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).

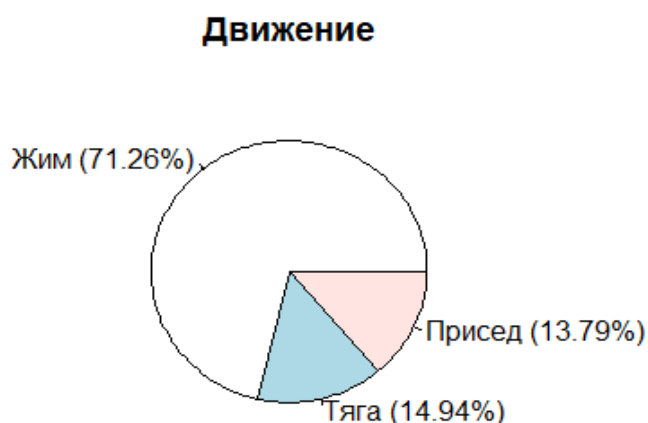


Рисунок 5

Распределение по другим факторам иллюстрируется на Рис. 6-11.

### Диапазон повторений

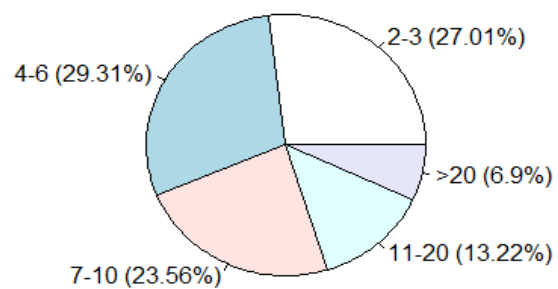


Рисунок 6

### Опыт тренировок

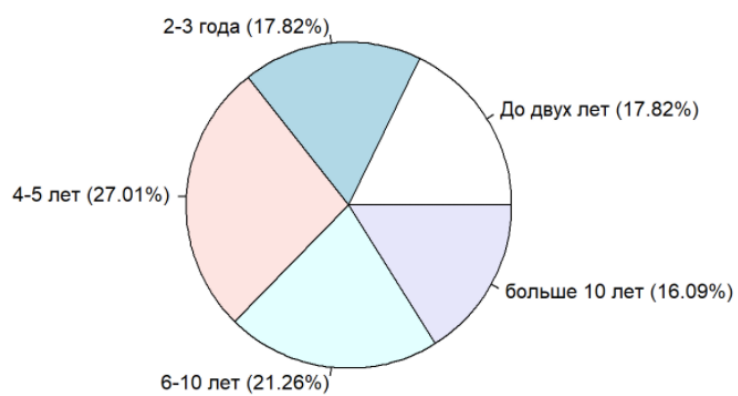


Рисунок 7<sup>11</sup>

### Возрастная группа

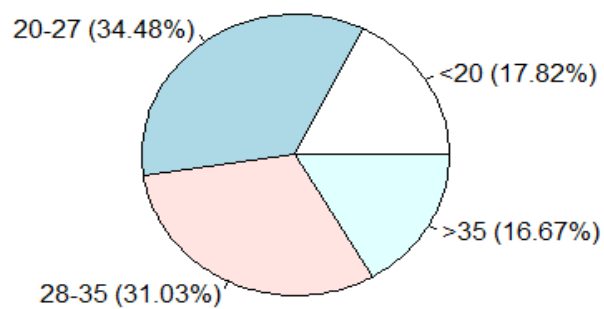


Рисунок 8

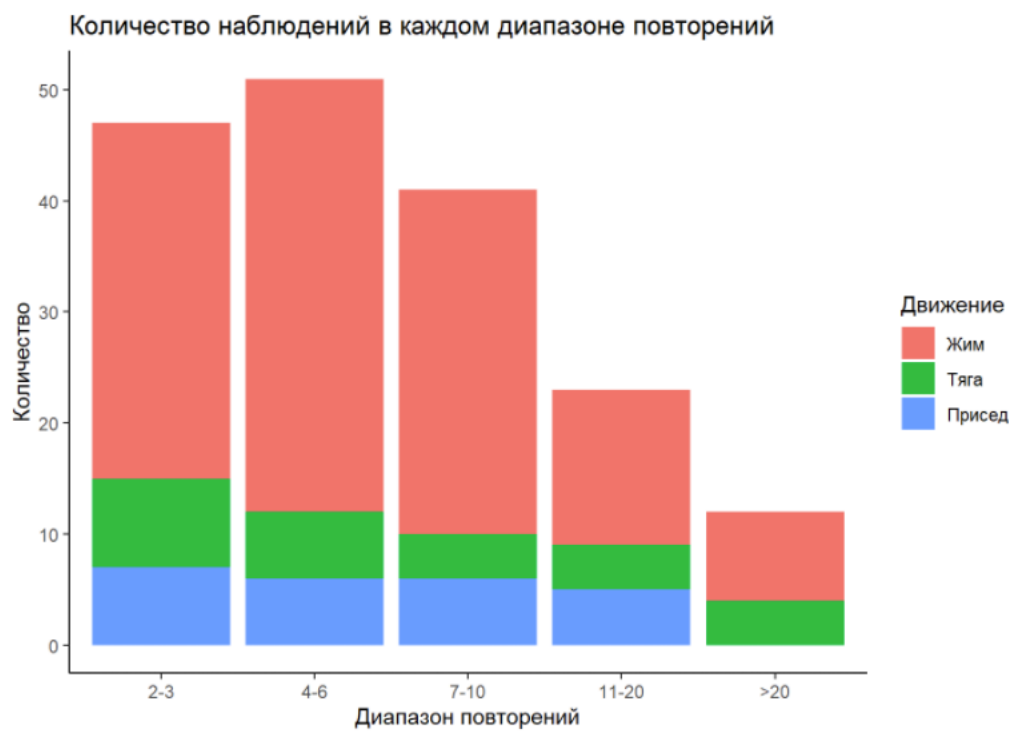


Рисунок 9

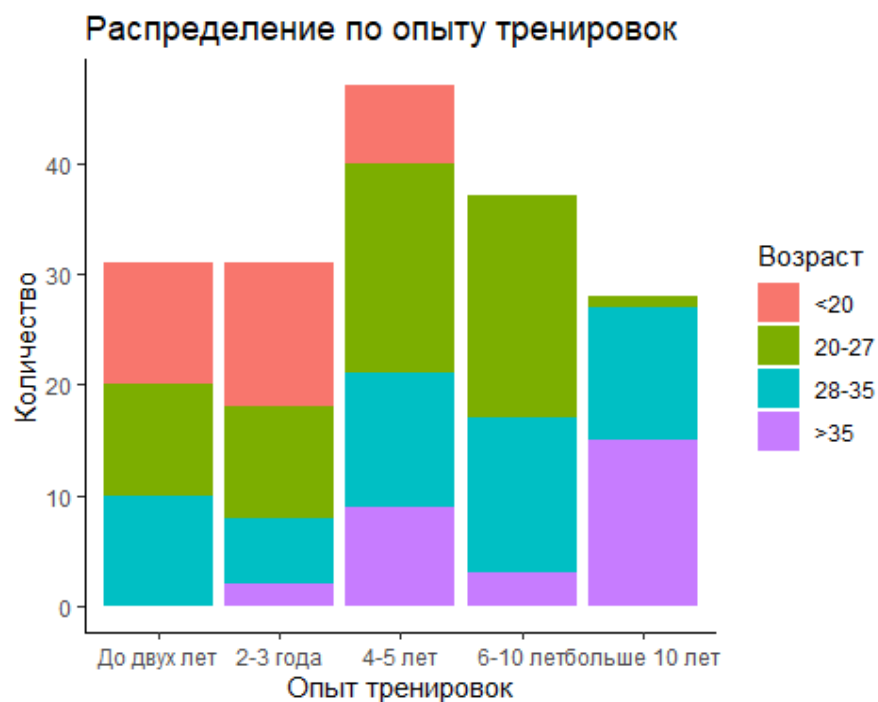
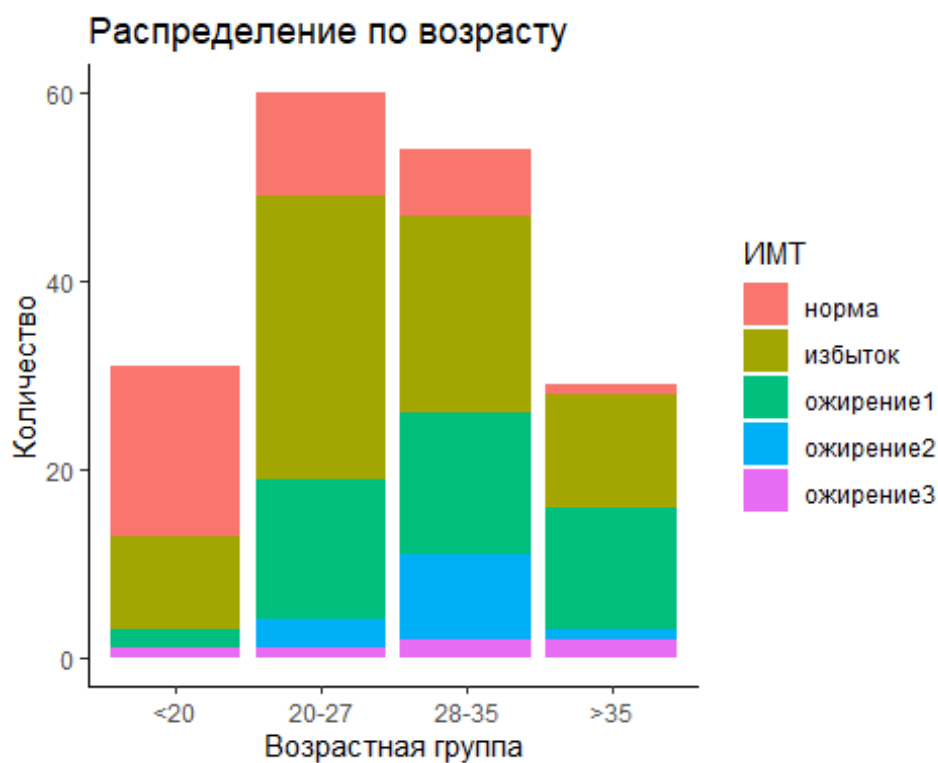


Рисунок 10



11  
Рисунок 11

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

	выраженный дефицит	дефицит	норма	избыток	ожирение1	ожирение2	ожирение3
Эктоморф	0	0	12	18	3	0	0
Мезоморф	0	0	19	23	5	1	0
Эндоморф	0	0	3	26	33	8	6

Рисунок 12

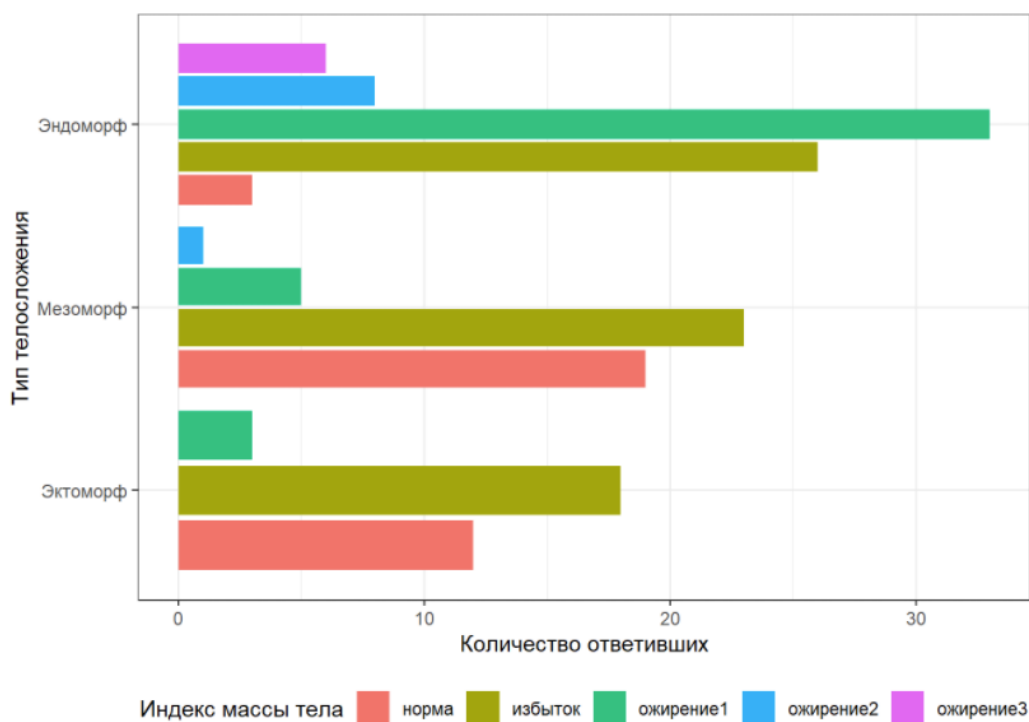


Рисунок 13

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 (Рис.14).

3-sample test for equality of proportions without continuity correction: (continued below)

Test statistic	df	P value	Alternative hypothesis	prop 1
22.88	2	1.073e-05 ***	two.sided	0.09091
prop 2	prop 3			
0.1042	0.4342			

2-sample test for equality of proportions with continuity correction:

Test statistic	df	P value	Alternative hypothesis	prop 1	prop 2
21.16	1	4.223e-06 ***	two.sided	0.09877	0.4342

Рисунок 14

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

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
BodyType	2	1231	615.6	33.82	4.201e-13
Residuals	171	3112	18.2	NA	NA

Welch Two Sample t-test: data %>% filter(BodyType != "Эндоморф") %>% select(Index) %>% and  
data %>% filter(BodyType == "Эндоморф") %>% select(Index) %>% t() %>% as.numeric() and t() %>% as.numeric() (continued below)

Test statistic	df	P value	Alternative hypothesis	mean of x
-8.105	152.2	1.608e-13 ***	two.sided	25.87
mean of y				
31.14				

Рисунок 15

## 2.2 Взаимодействия переменных

Корреляции количественных переменных приведены на Рис. 16.

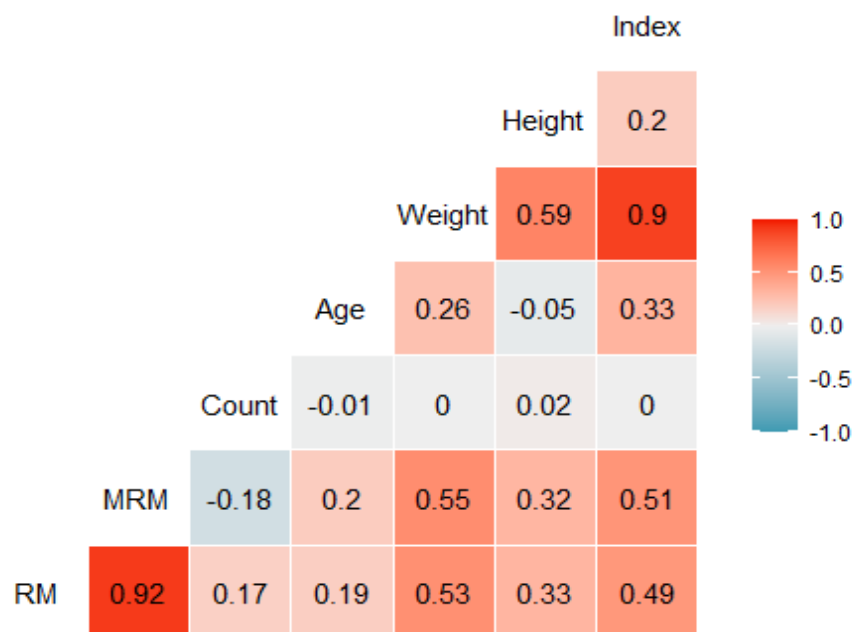


Рисунок 16

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

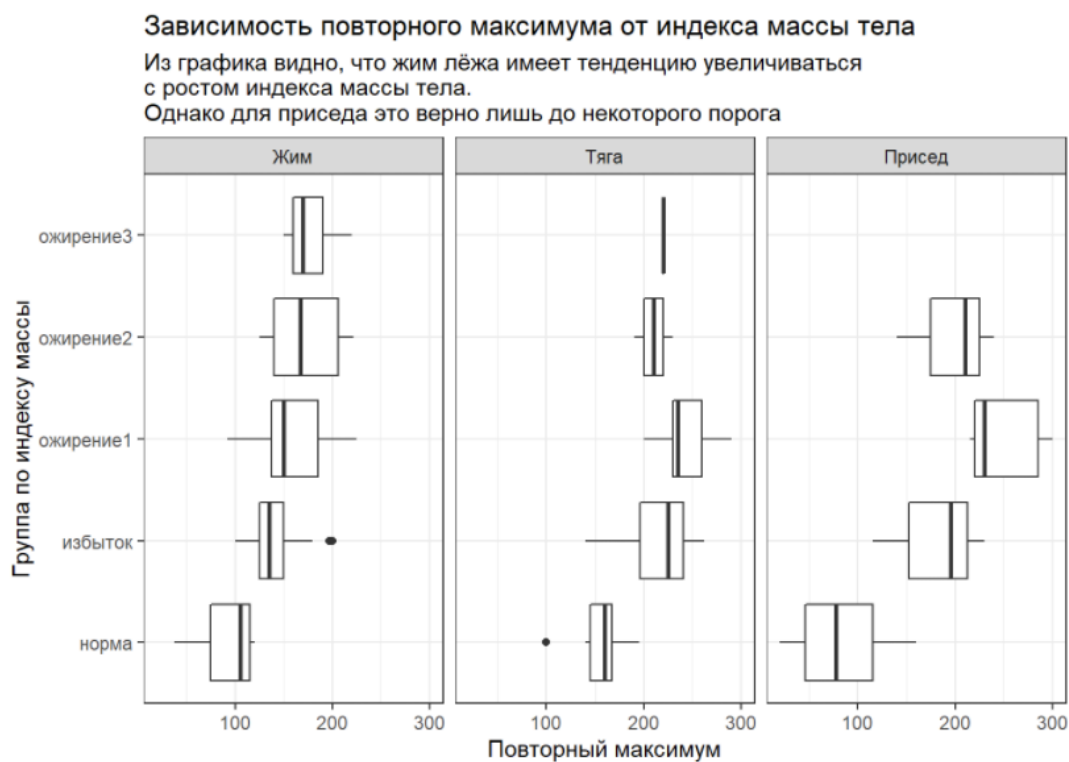


Рисунок 17



Рисунок 18

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



### Оценка параметра для разного числа повторений

Красным цветом обозначен параметр 0.0333 из книги Лилли.  
Как видно, он может быть верен для числа повторений от 2 до 5

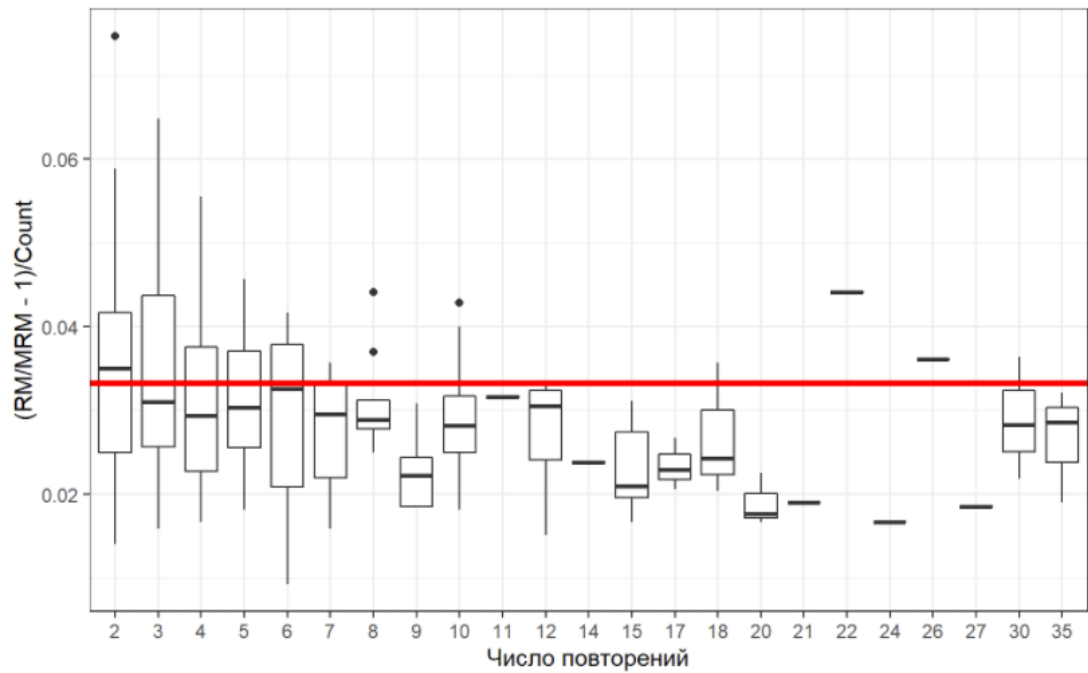


Рисунок 19

Из следующего графика (Рис. 20) может следовать, что начиная с 8-9 повторений зависимость между МПМ и ПМ ослабевает.

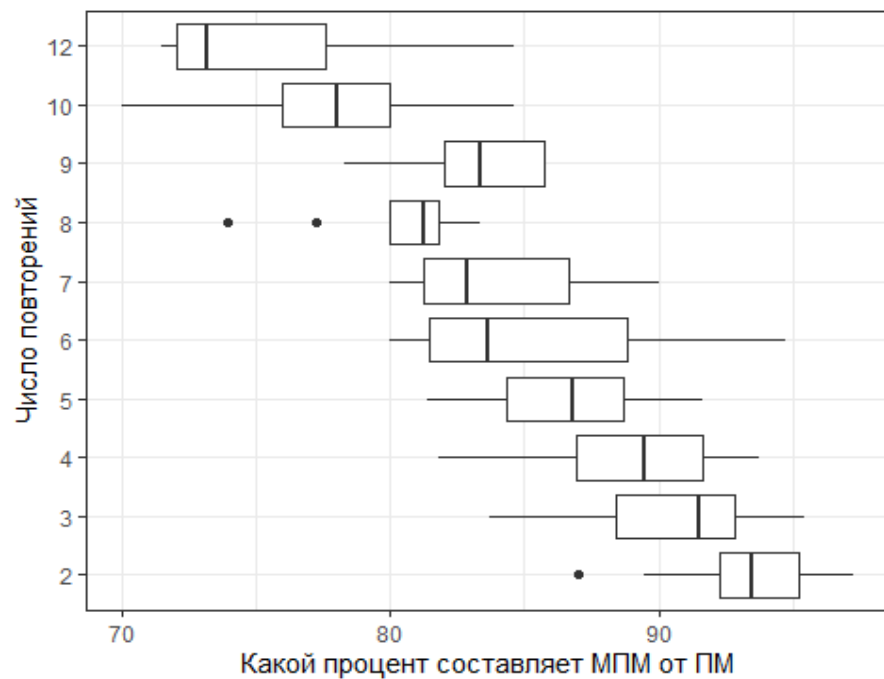


Рисунок 20

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

Корреляция между MRM & RM для каждого повторения

Число повторений	Корреляция MRM & RM	Нижняя граница	Верхняя граница	Всего наблюдений
2	0.9980117	0.9949986	0.9992103	21
3	0.9919420	0.9818449	0.9964336	26
4	0.9947569	0.9838325	0.9983060	15
5	0.9891096	0.9734436	0.9955548	22
6	0.9942644	0.9814191	0.9982374	14
7	0.9936119	0.9687446	0.9987074	9
8	0.9964144	0.9876689	0.9989606	13
9	0.9581086	0.4902757	0.9973273	5
10	0.9890079	0.9645992	0.9966159	14

Рисунок 21

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

р-значения для разных факторных переменных на каждом числе повторений

Число повторений	р-значения для телосложений	р-значения для типа движения	р-значения для групп по индексу массы
2	0.9458770	0.7730952	0.8959581
3	0.8956176	0.1384093	0.2062476
4	0.9985364	0.279419	0.8345974
5	0.3427142	0.9919357	0.5228145
6	0.7269629	0.2637337	0.3154734
7	0.8204683	мало данных	0.9085377
8	0.6248755	мало данных	0.7177698
9	0.2564939	мало данных	0.4490970
10	0.0878330	мало данных	0.7676728

Рисунок 22

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

Какой % составляет МПМ от ПМ

Число повторений	Ожидаемый %	Нижняя граница	Верхняя граница
2	93.54267	92.03110	95.05425
4	89.07451	86.50007	91.64895
6	85.58991	81.56746	89.61236
8	80.55848	78.43606	82.68090
10	77.84473	74.63339	81.05607
12	75.48574	68.66088	82.31060
15	74.47497	64.50324	84.44670
20	72.62619	54.19540	91.05698

Рисунок 23

В целом, эти данные согласуются с тем, что используются National Strength and Conditioning Association (NSCA)<sup>21</sup> (Рис. 24).

% of 1RM	Number of Repetitions That Can Be Performed
100	1
95	2
90	4
85	6
80	8
75	10
67	12
65	15

Рисунок 24

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%.

---

<sup>21</sup> <https://www.nsca.com>

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

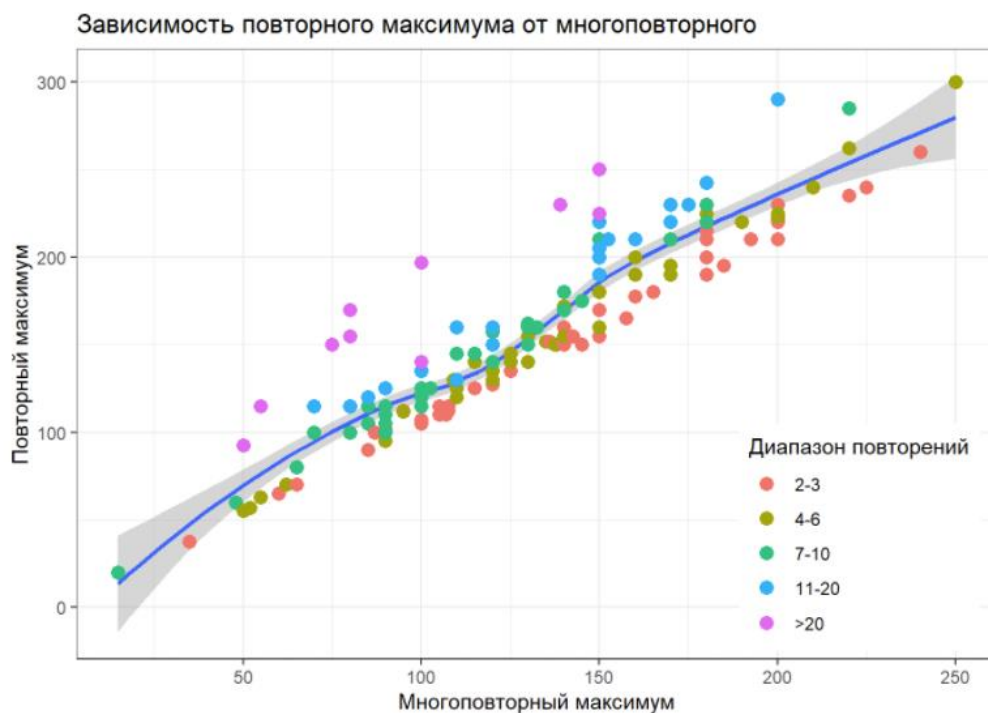


Рисунок 25

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. Рис. 26.

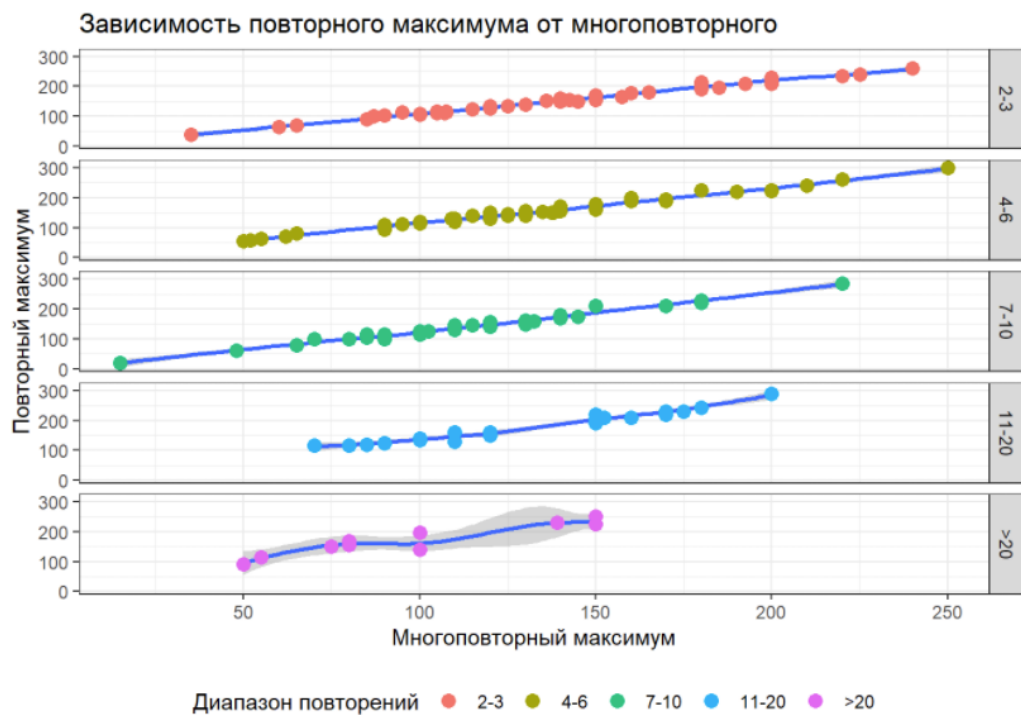


Рисунок 26

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

Есть ли зависимость между отношением  $\frac{RM}{MRM}$  в зависимости от опыта тренировок и других факторов? Дисперсионный анализ показывает, что нет (Рис. 27).

Значимость различий отношения ПМ к МПМ в каждом диапазоне повторений при разных уровнях факторных переменных

Диапазон повторений	Опыт	Возраст	ИМТ	Телосложение
2-3	0.6506219	0.7671962	0.5516303	0.7779869
4-6	0.6178563	0.6907457	0.6655732	0.5432097
7-10	0.1237067	0.8407336	0.8969795	0.3628399
11-20	0.7293524	0.0354352	0.0304071	0.6845562
>20	0.8249503	0.5901704	0.2511299	0.3736225

Рисунок 27

Единственное: обнаружилась разница в зависимости от возраста и индекса массы тела для диапазона повторений 11-20 (Рис. 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.

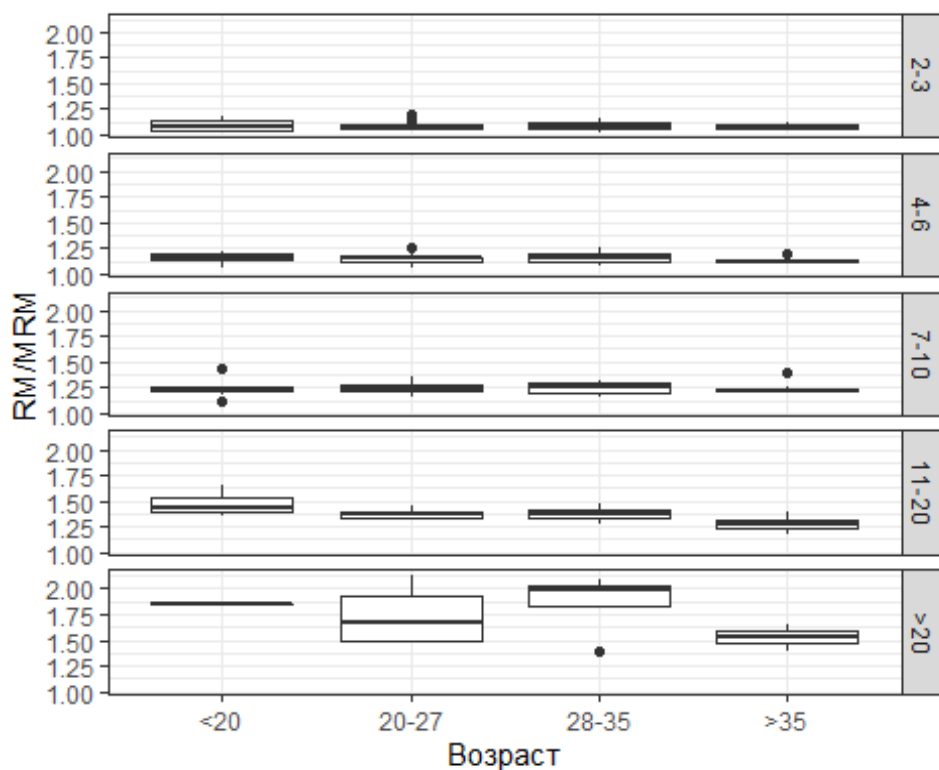


Рисунок 28

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

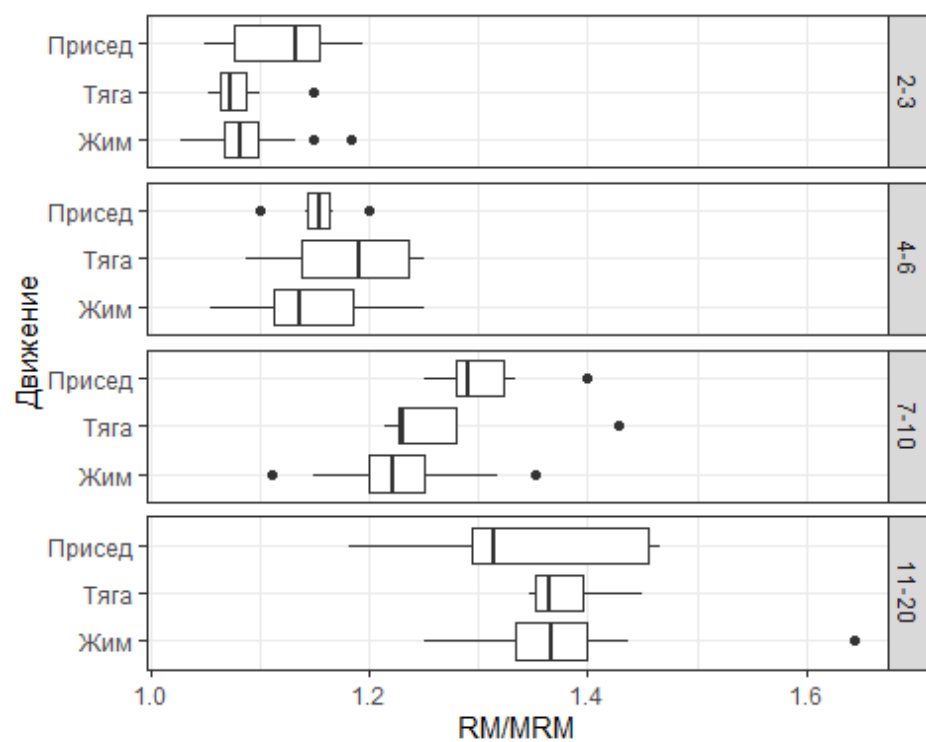


Рисунок 29

Дополнительные закономерности изображены на Рис. 30-32.

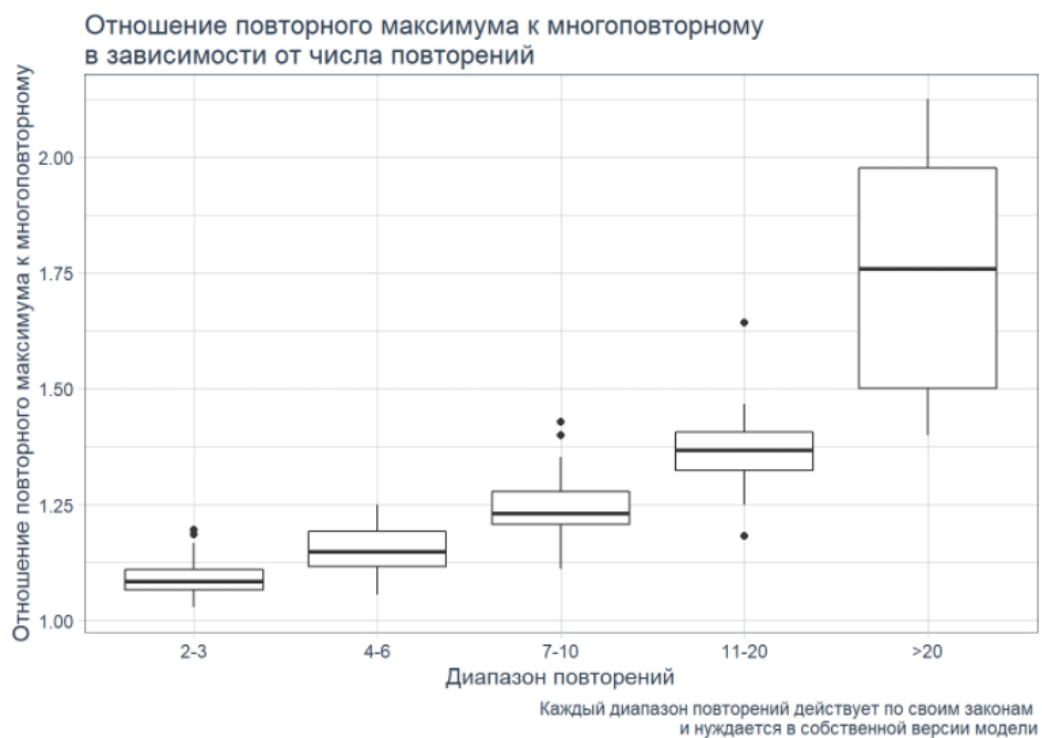


Рисунок 30

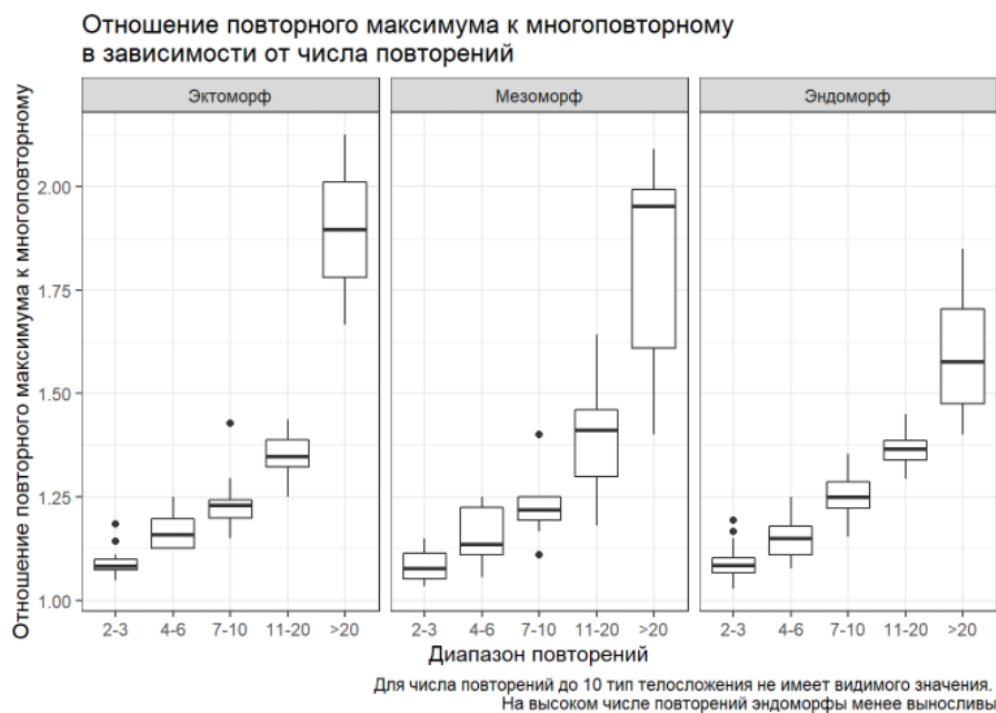


Рисунок 31

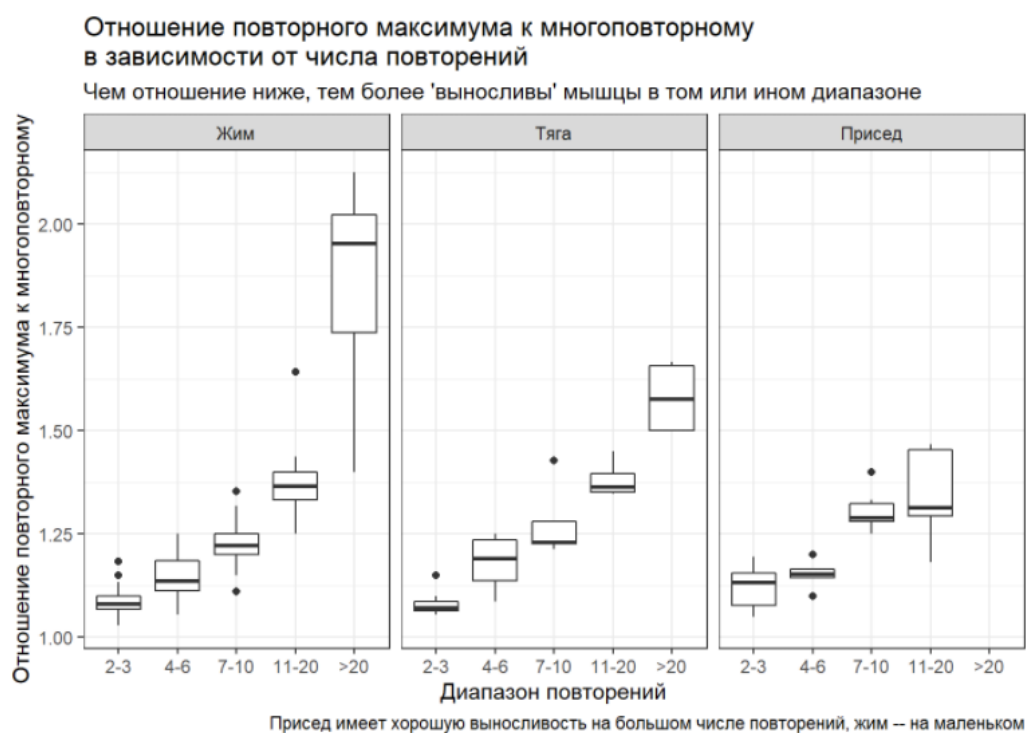


Рисунок 32



## 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 прохождении использовал экипировку или другую технику упражнения, либо, опять же, опечатался; в итоге жим 170x5 оказывался равным 220, хотя ясно, что тут должны стоять 170x5 и 200 либо 170x15 и 220.

С такими *выбросами* можно сделать только две вещи:

- 1) исправить их, опираясь на логику либо переспросить того, кому это наблюдение принадлежит<sup>22</sup>;
- 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<sup>23</sup> содержатся комментарии, где и что я исправил и почему.

---

<sup>22</sup> к сожалению, этого не получилось сделать, так как испытуемые редко пользуются электронной почтой, видимо

<sup>23</sup> <https://github.com/PasaOpasen/Powerlifting-training-diary-and-articles/blob/master/Estimating%20RM/data.tsv>

## 2.4 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  $120 * 3 = 140$ , 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<sup>24</sup>, 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<sup>25</sup>; вдобавок, для таких предсказаний сложнее определить аномальные наблюдения в смысле предыдущего пункта; кроме

---

<sup>24</sup> 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

<sup>25</sup> такую нижнюю границу можно предсказать, просто умножив вес примерно на 1.5, исследование же рассчитано на поиск верхней границы

того, даже на графиках видно, что диапазон выше 20 очень сильно отличается от остальных диапазонов.

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

RM		MRM		Count		Action		Sex	
Min.	: 20.0	Min.	: 15	Min.	: 2.000	Жим	:116	Мужчина:	153
1st Qu.:	120.0	1st Qu.:	100	1st Qu.:	3.000	Тяга	: 22	Женщина:	9
Median	:145.0	Median	:125	Median	: 5.000	Присед:	24		
Mean	:152.6	Mean	:129	Mean	: 6.636				
3rd Qu.:	190.0	3rd Qu.:	150	3rd Qu.:	9.000				
Max.	:300.0	Max.	:250	Max.	:20.000				

Experience		Age		Weight		Height		BodyType	
До двух лет	:31	Min.	:15	Min.	: 47.00	Min.	:155	Эктоморф:	34
2-3 года	:28	1st Qu.:	21	1st Qu.:	78.00	1st Qu.:	170	Мезоморф:	46
4-5 лет	:45	Median	:27	Median	: 87.00	Median	:177	Эндоморф:	82
6-10 лет	:32	Mean	:28	Mean	: 89.59	Mean	:177		
больше 10 лет:	26	3rd Qu.:	33	3rd Qu.:	102.75	3rd Qu.:	183		
		Max.	:62	Max.	:160.00	Max.	:198		

CountGroup	AgeGroup	Index		IndexGroup			
2-3	:47	<20	:30	Min.	:18.83	выраженный дефицит:	0
4-6	:51	20-27:	55	1st Qu.:	25.17	дефицит	: 0
7-10	:41	28-35:	50	Median	:27.76	норма	:36
11-20:	23	>35	:27	Mean	:28.44	избыток	:66
>20	: 0			3rd Qu.:	31.39	ожирение1	:42
				Max.	:45.27	ожирение2	:12
						ожирение3	: 6

Рисунок 33

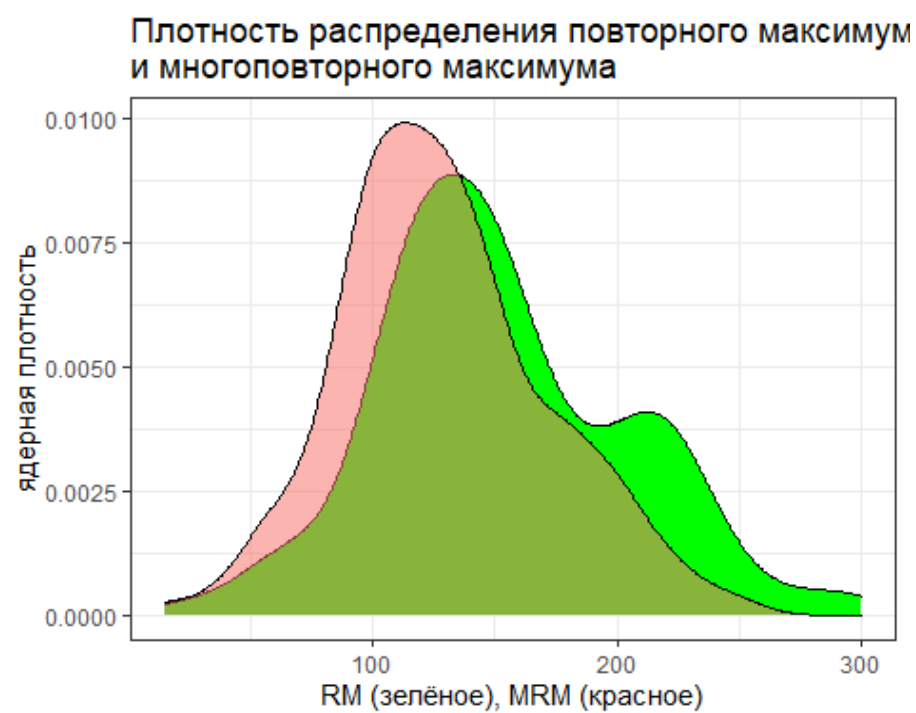


Рисунок 34

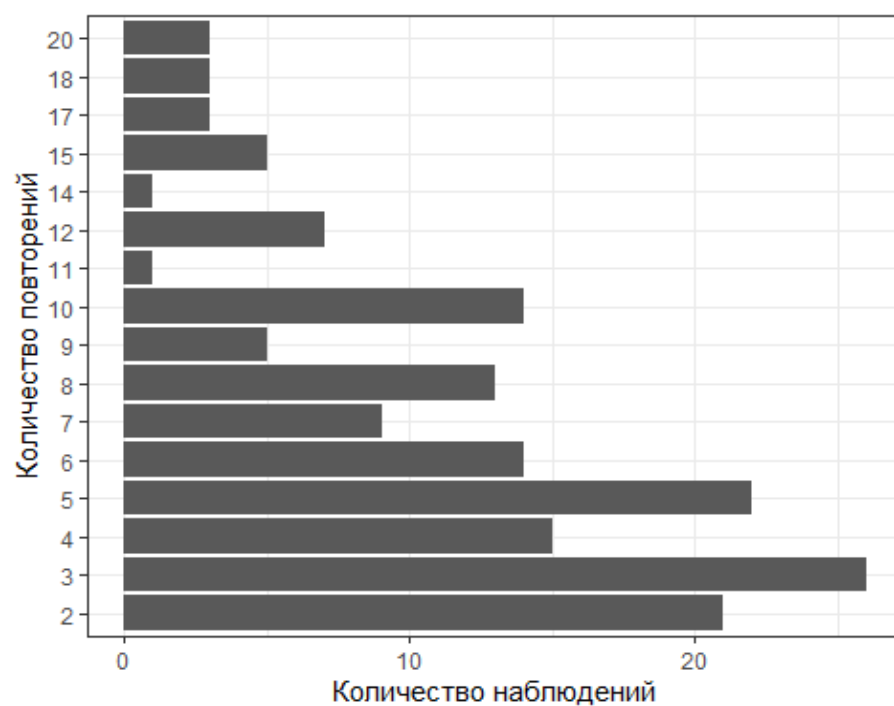


Рисунок 35

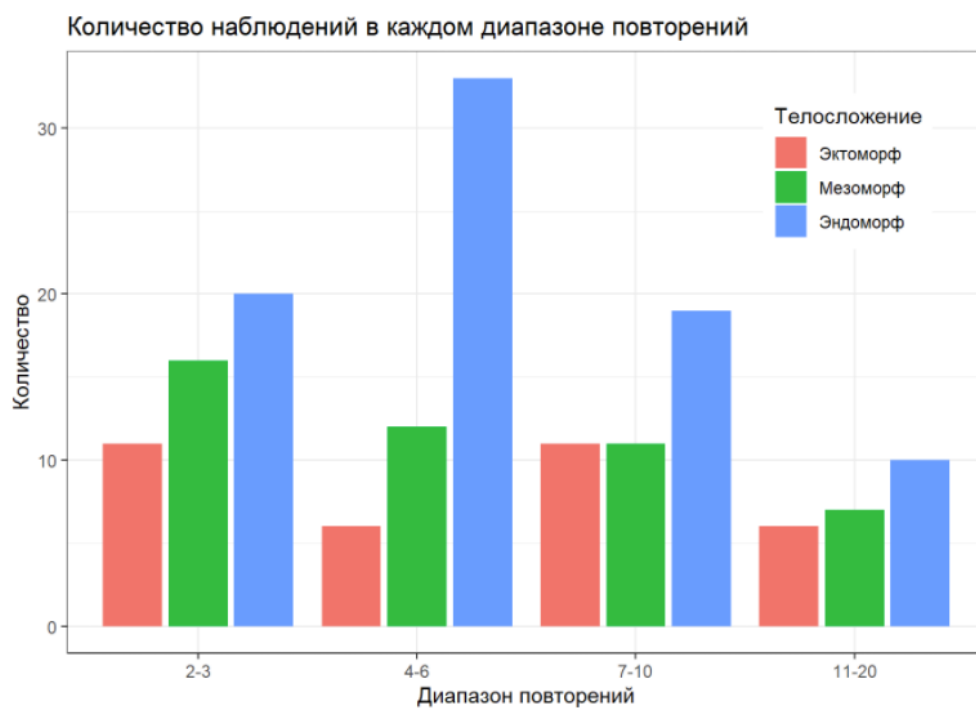


Рисунок 36

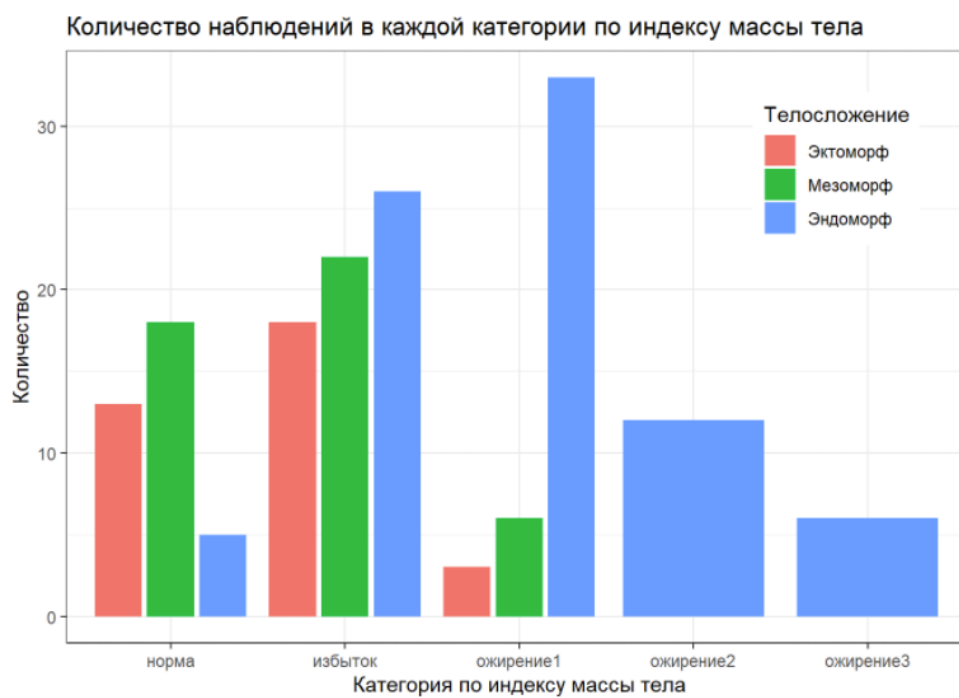


Рисунок 37

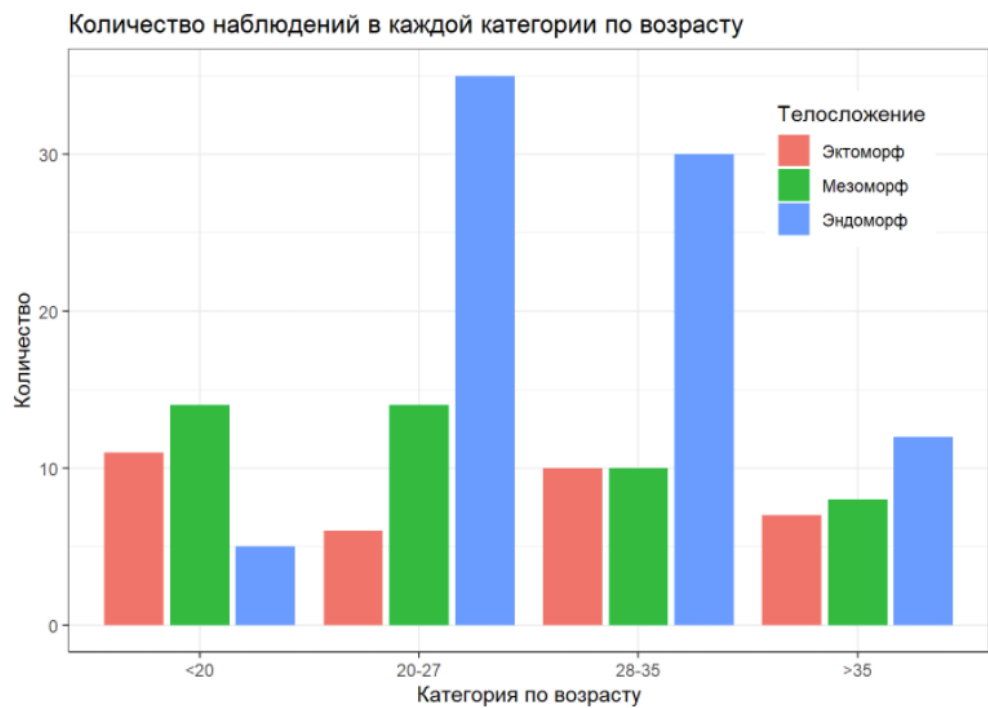


Рисунок 38

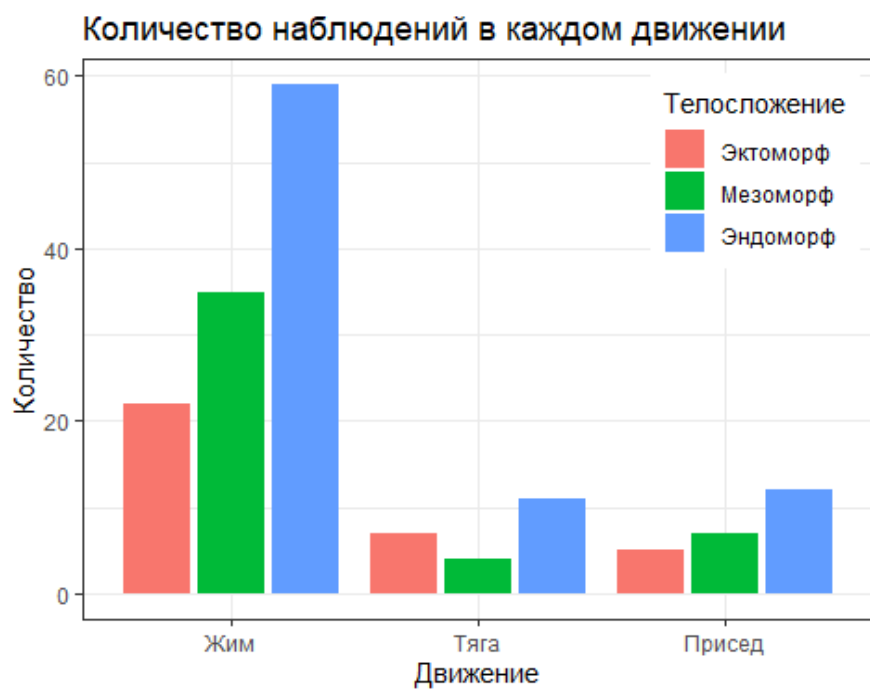


Рисунок 39

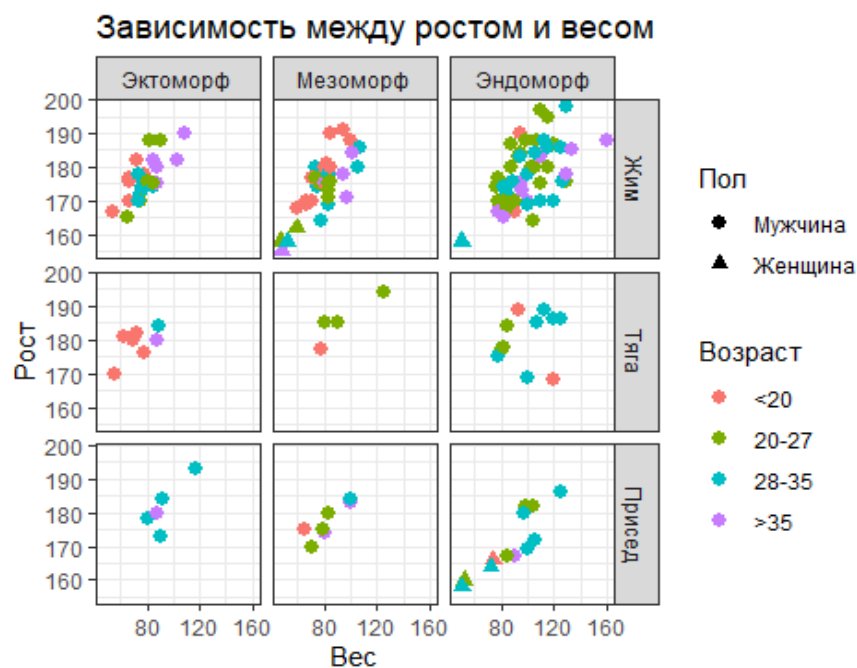


Рисунок 40

## 2.5 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 мезоморфов – в жиме, у эктоморфов – в тяге, у эндоморфов – в приседе;

5) Не обнаружено значимой разницы значений  $\frac{MRM}{RM}$  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); процентки 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)  $RM$  (повторный максимум) обязательно зависит как от  $MRM$  (многоповторный максимум) и  $Count$  (числа повторений), так и от  $CountGroup$  – группы по числу повторений;

2)  $RM$  может зависеть также от  $BodyType$  (типа телосложения) и  $Action$  (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  $RM$  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  $RM$  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  $RM = MRM \cdot (1 + 0.0333 \cdot Count)$  ко всем наблюдениям, получим оценки  $RM$  (Fact), которые будут отличаться от истинного  $RM$  (Target). Разница между этими величинами – остатки. Для модели Вендлера остатки приведены на Рис. 41-43.

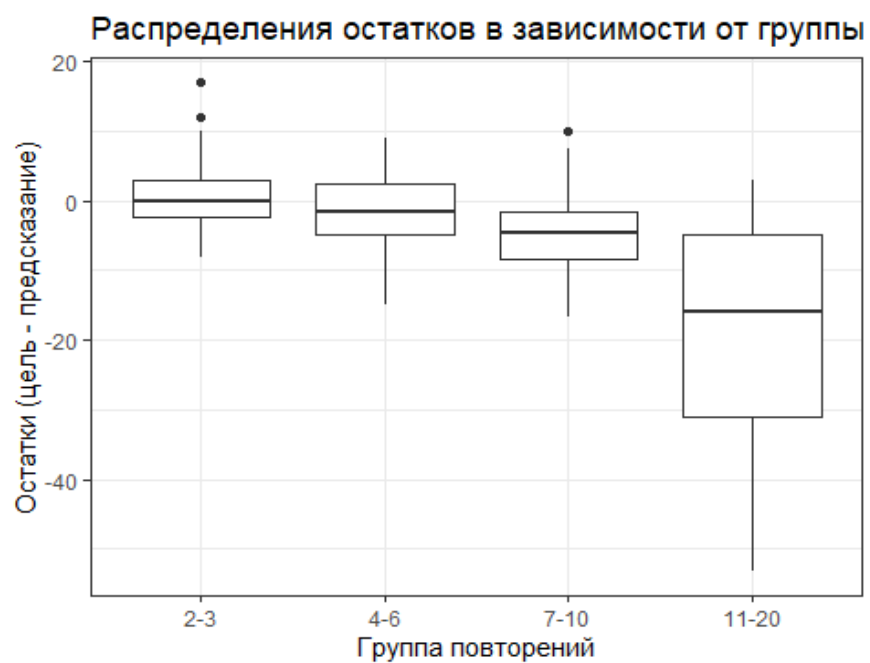


Рисунок 41

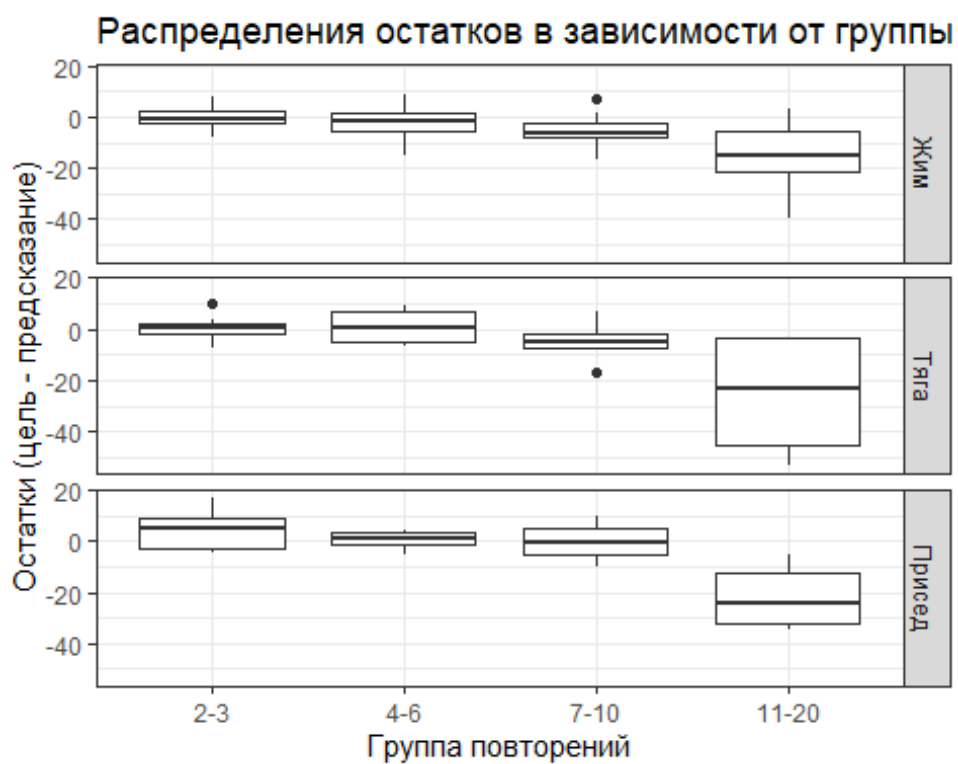


Рисунок 42



Рисунок 43

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

$$\log\left(\frac{RM}{MRM}\right) = a + b \cdot Count, \quad (4)$$

где  $a, b$  зависят от того, к какой части тела относится упражнение. Сама идея сделать логарифмическое преобразование используется нередко, но здесь это имеет негативный эффект:  $R^2$  равен далеко не 0.99, как сказано на сайте (где я нашёл эти формулы), реальный  $R^2$  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  $RM = MRM \cdot (a + b \cdot Count)$  и подберём коэффициенты  $a, b$  лучшим образом. В итоге получим результаты, изображенные на Рис. 44-49.

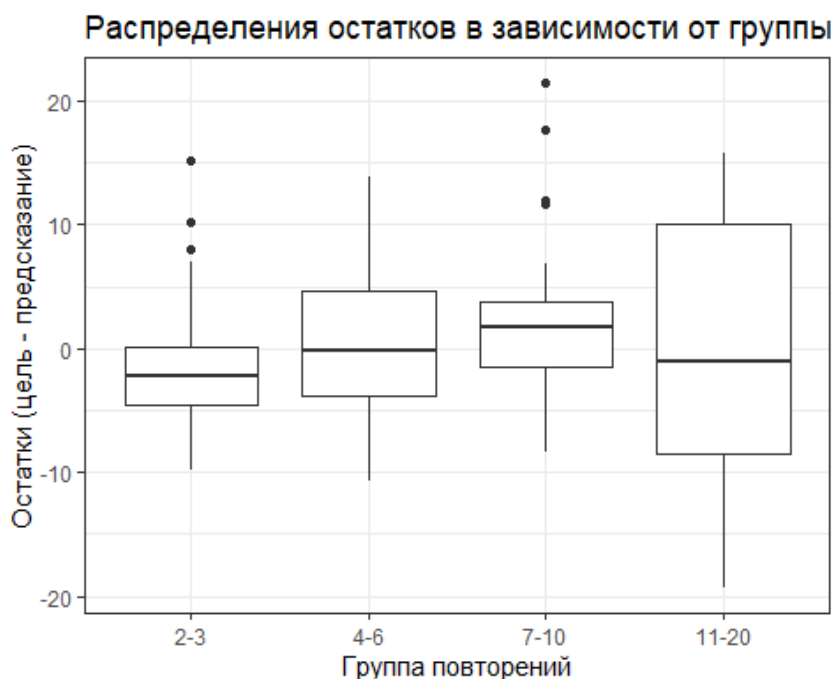


Рисунок 44 11

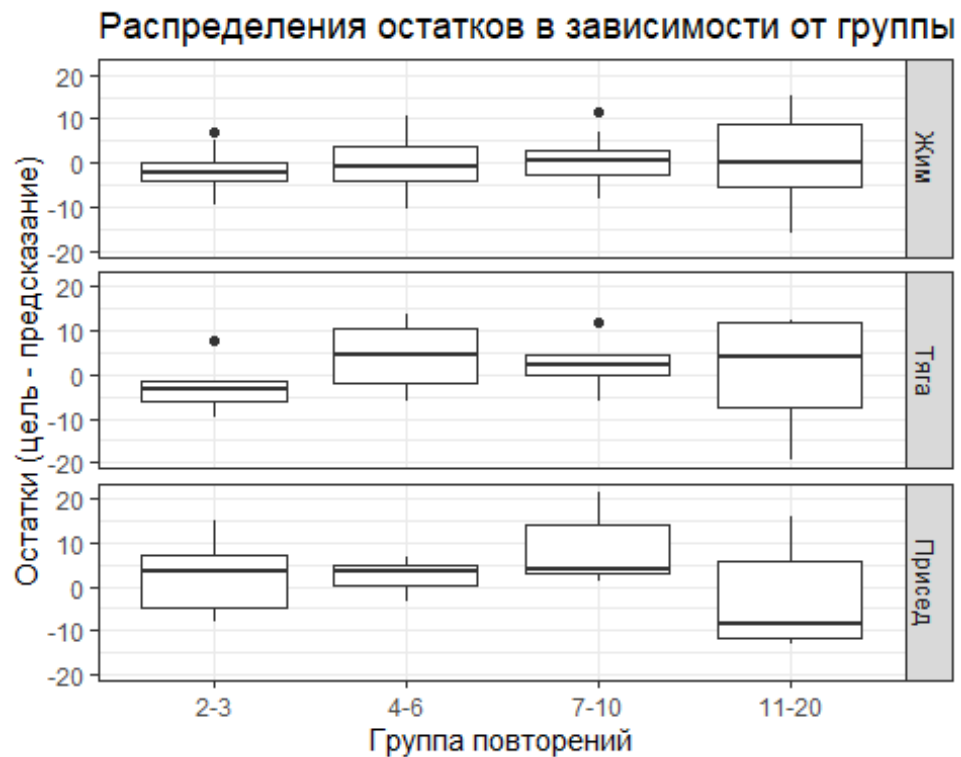


Рисунок 45

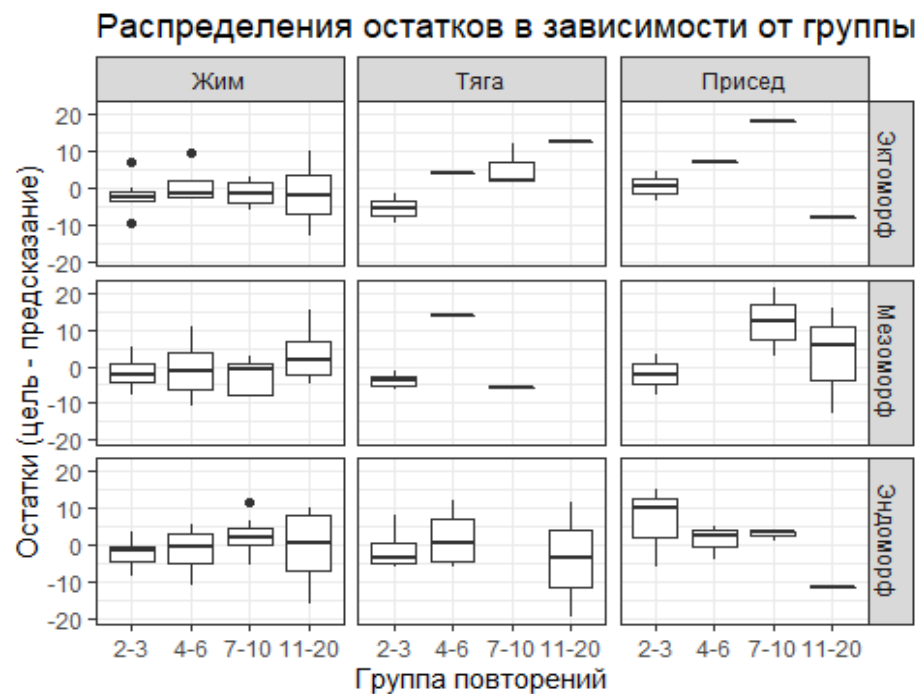


Рисунок 46

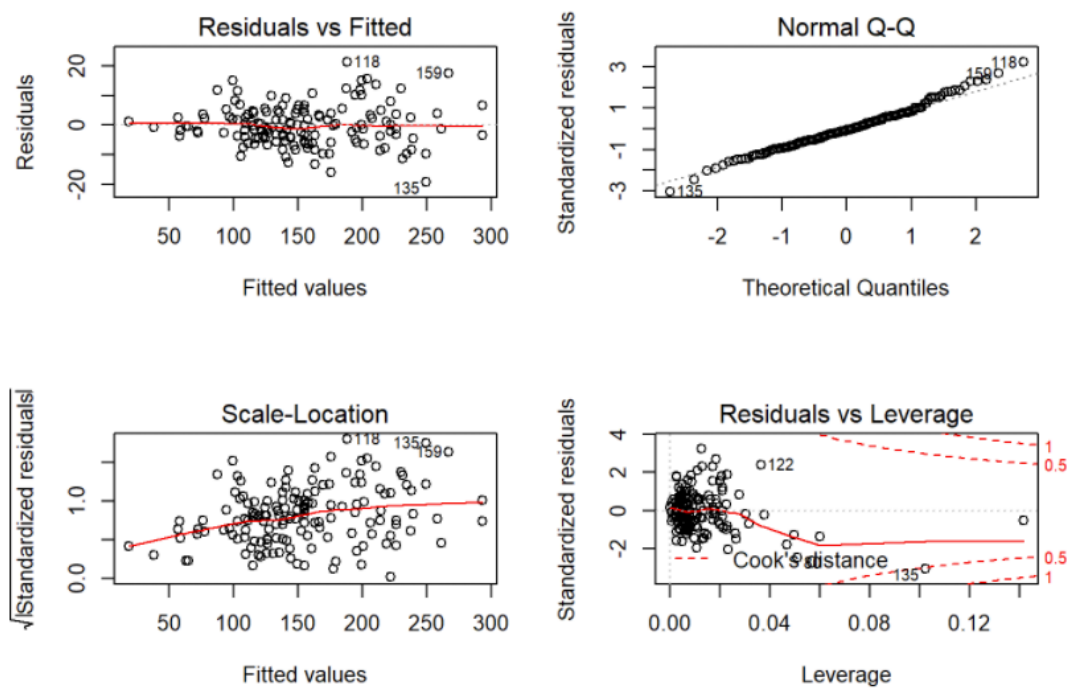


Рисунок 47

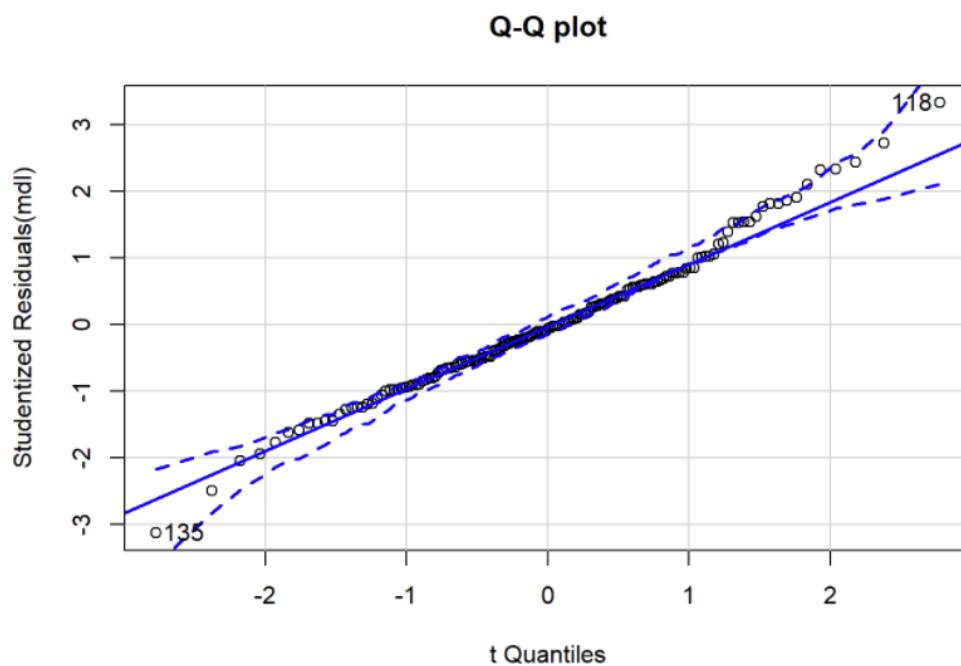
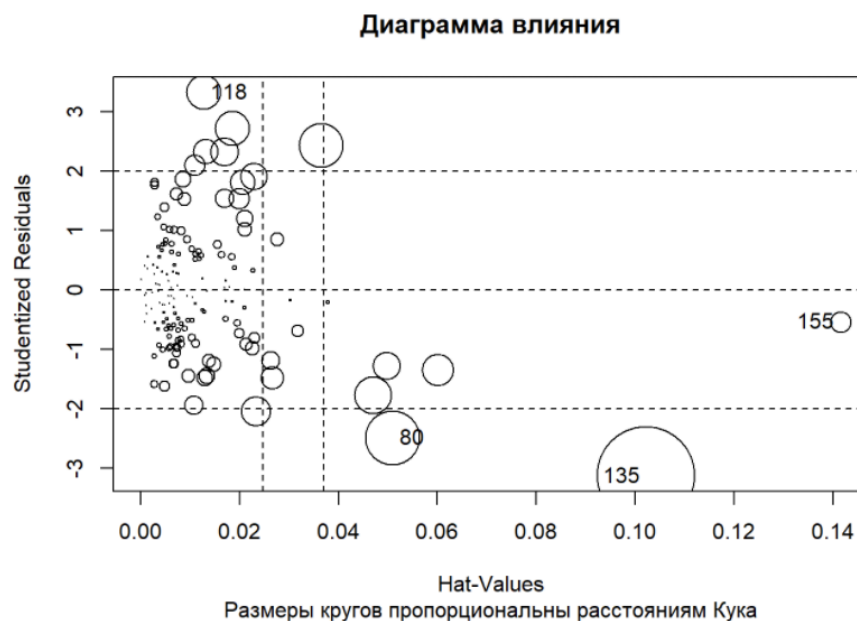


Рисунок 48



11  
**Рисунок 49**

Далее приведена вся необходимая информация о качестве модели (назовём эту модель  $b_1$ ):

- 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  $a, b$  в зависимости от факторных переменных.

Путём подбора удалось найти две схожие модели (назовём их  $b_2$  и  $b_3$ ), 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  $RM = MRM \cdot (a + b \cdot Count)$ , в которой и  $a$  и  $b$  имеют поправку на диапазон повторений. Результаты представлены на Рис. 50-54.

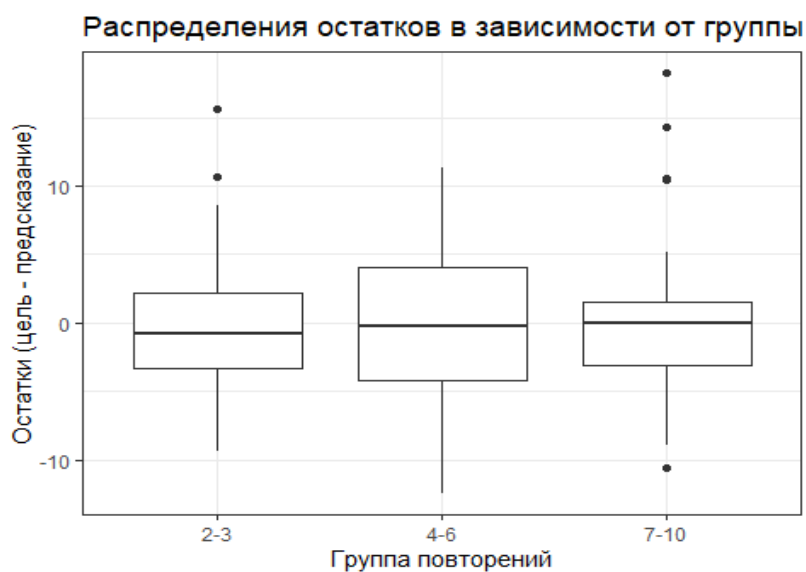


Рисунок 50 <sup>11</sup>

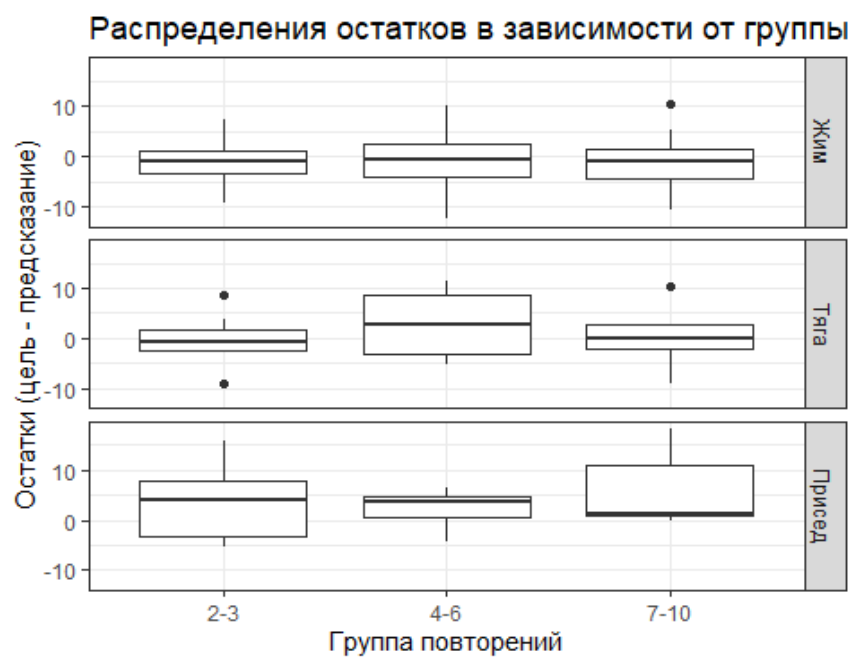


Рисунок 51

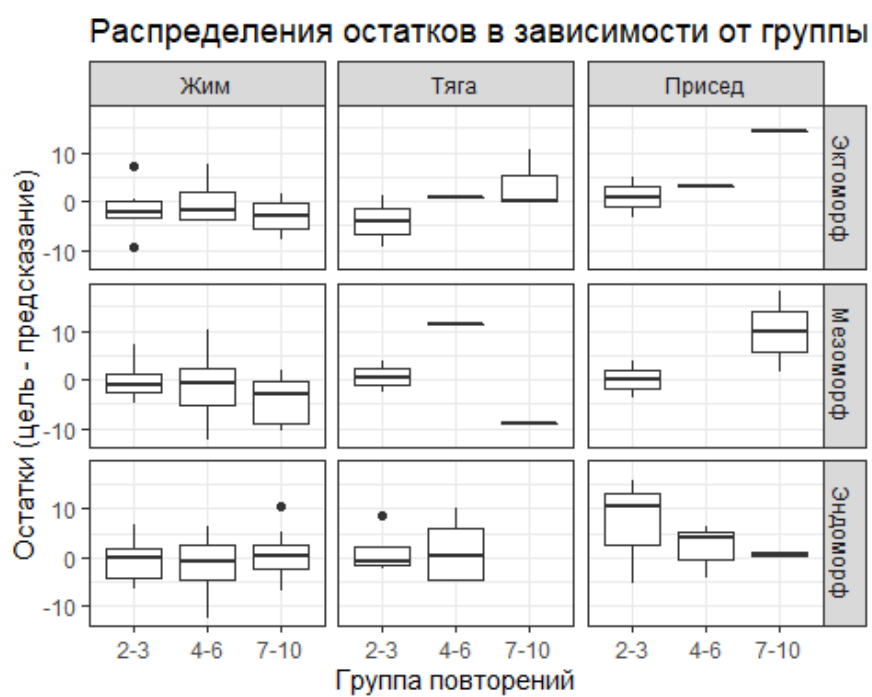


Рисунок 52<sup>11</sup>

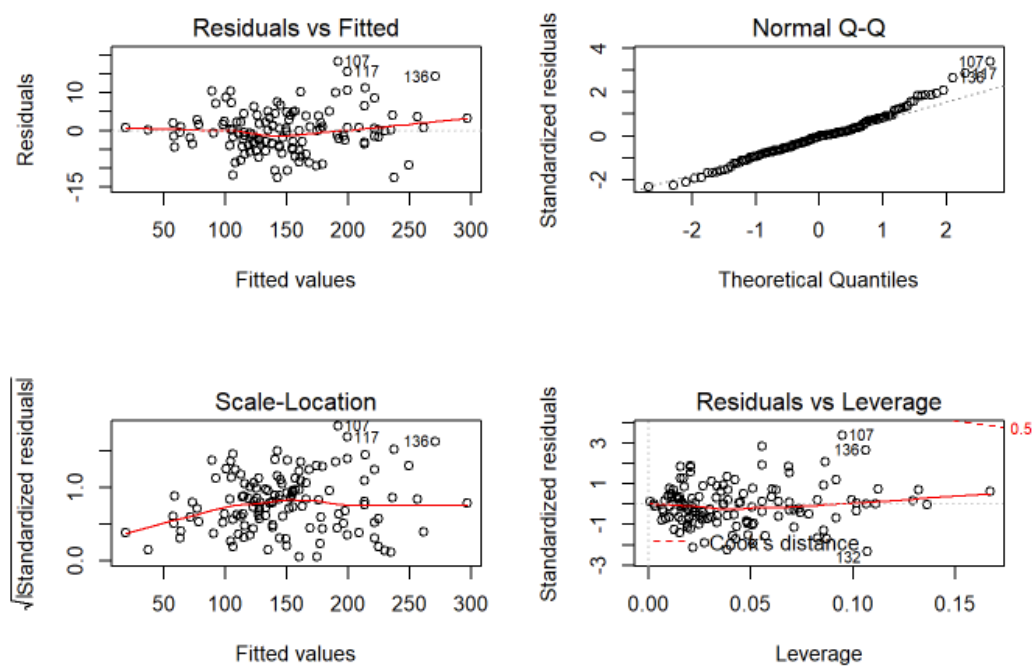


Рисунок 53

### Диаграмма влияния

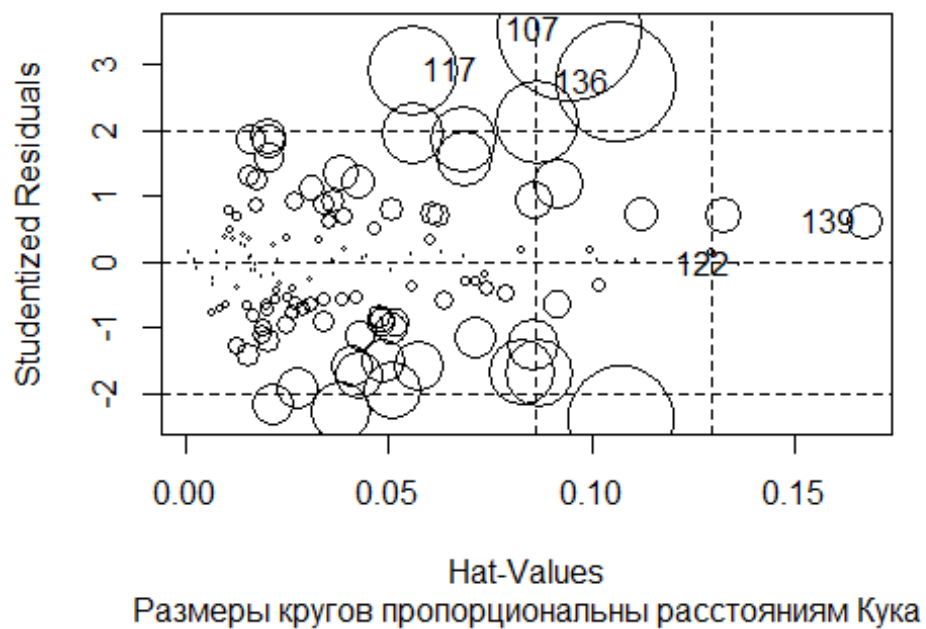


Рисунок 54

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 ( $b_4$  и  $b_5$ ), которые, несмотря на большую сложность в сравнении с предыдущими, имеют лучшие оценки при кросс-валидации.

Статистика для модели  $b_5$  приведена на Рис. 55-58.

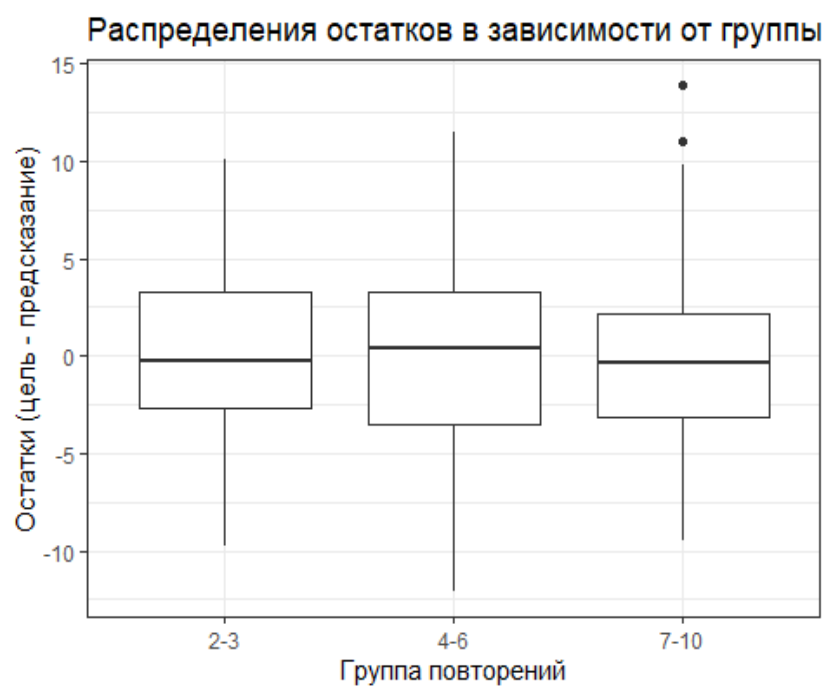


Рисунок 55

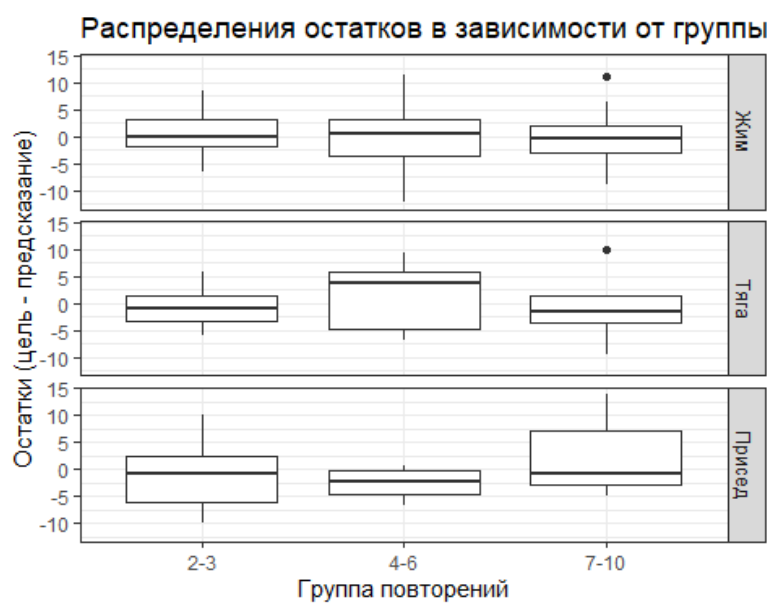


Рисунок 56

Распределения остатков в зависимости от группы

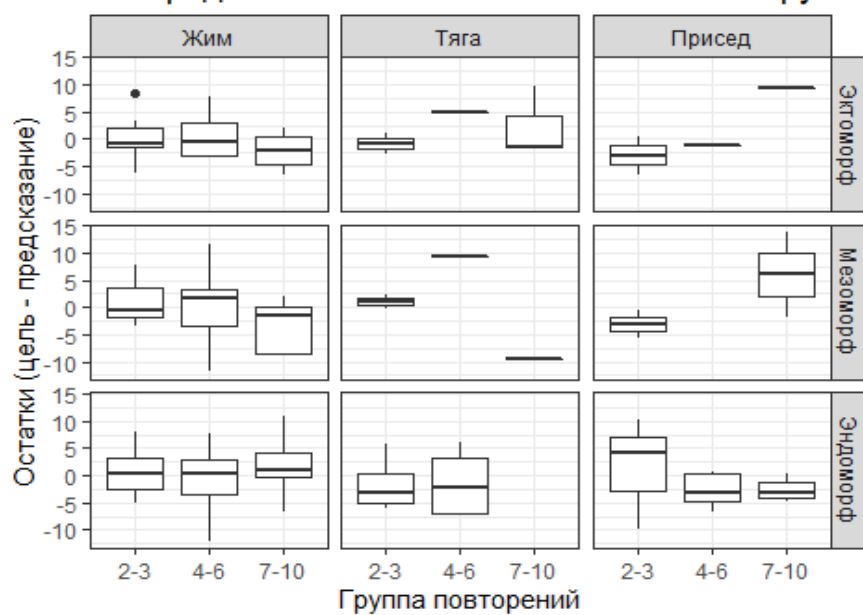


Рисунок 57

Диаграмма влияния

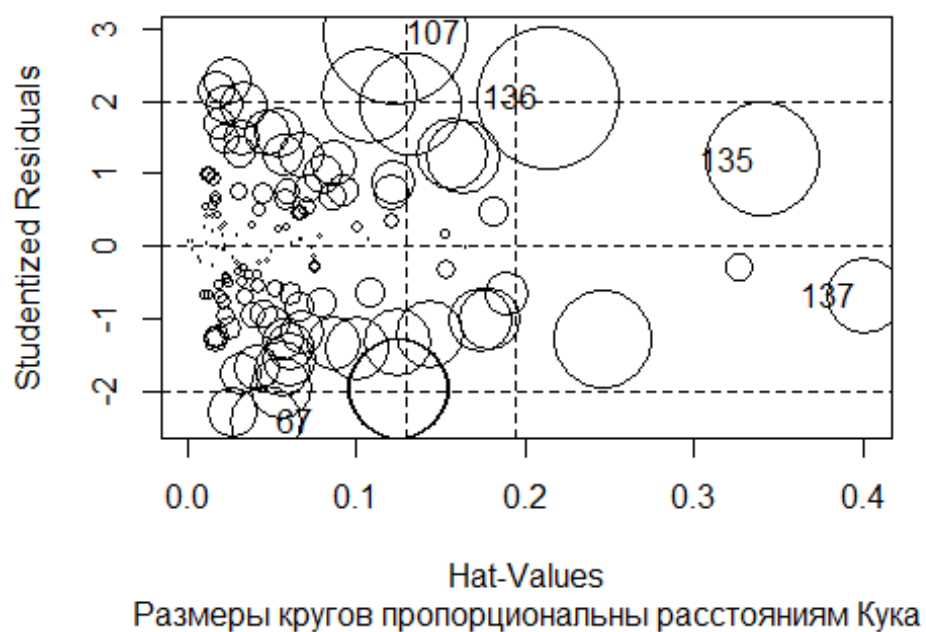


Рисунок 58

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  $b_5$  значительно превосходит остальные модели, причём это верно, если делать проверку как на всём диапазоне повторений (2-10), так и на более близком к силовому диапазону 2-7, см. Рис. 59.

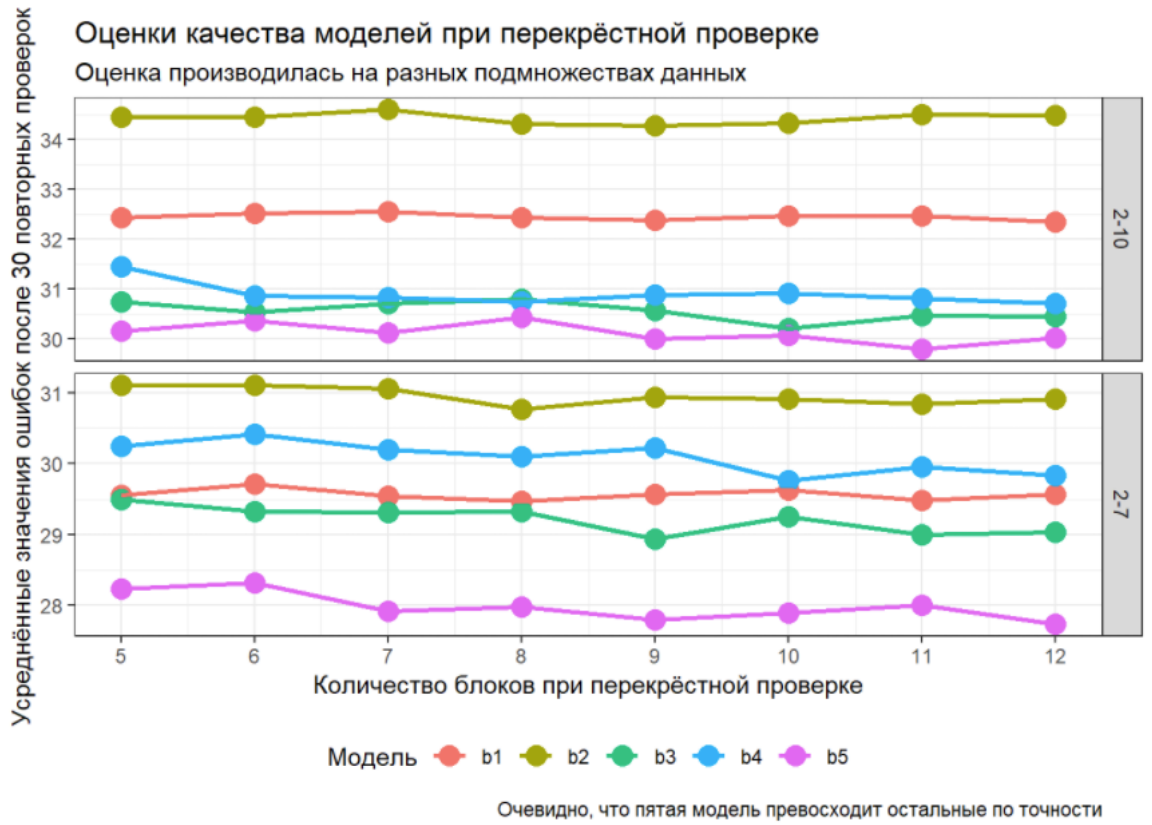


Рисунок 59

На втором месте после модели  $b_5$  идёт более простая модель  $b_3$ .

Что это за модели? Посмотрим на модель  $b_3$ . Она имеет вид

$$RM = MRM \cdot (Action_{coef} + CountGroup_{coef} \cdot Count), \quad (5)$$

где  $Action_{coef}$  равен 0.9988949 для жима, 1.0129181 для тяги и 1.0362281 для приседа, а  $CountGroup_{coef}$  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 модель  $b_3$  превосходит 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 Рис. 60.

На сколько нужно умножить свой вес для разного числа повторений, чтобы получить 1ПМ

Повторы	Присед	Жим	Тяга
1	1.000000	1.000000	1.000000
2	1.098673	1.061340	1.075363
3	1.129895	1.092562	1.106585
4	1.154553	1.117220	1.131243
5	1.184134	1.146801	1.160824
6	1.213715	1.176382	1.190405
7	1.227700	1.190367	1.204390
8	1.255053	1.217720	1.231743
9	1.282406	1.245073	1.259096
10	1.309759	1.272426	1.286449

Рисунок 60

Модель  $b_5$  имеет два дополнительных слагаемых. Её вид, аналогично:

$$RM = MRM \cdot (CountGroup_{coef_2} + Action_{coef} + CountGroup_{coef_1} \cdot Count) + coef \cdot \left( \frac{MRM}{Index} \right)^6 \quad (6)$$

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

$$\frac{MRM}{Index} = \frac{MRM \cdot (0.01 \cdot Height)^2}{Weight} = \frac{MRM \cdot Height^2}{10000 \cdot Weight} \quad (7)$$

Вообще существование этого слагаемого (с шестой степенью) кажется чем-то очень сомнительным, но математика показывает, что это имеет смысл. Результаты дисперсионного анализа (Рис. 60) говорят о том, что включение указанного коэффициента создаёт значимое различие между моделями с ним и без него:

Analysis of Variance Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
131	3703	NA	NA	NA	NA
130	3512	1	190.7	7.059	0.008874

Рисунок 61

Очень интересно, что в этих двух моделях нет зависимости от возраста, типа телосложения, опыта и пола<sup>26</sup>. 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.

---

<sup>26</sup> по крайней мере, наблюдения от женщин, прошедших опрос, не становились выбросами для этих моделей, а хорошо вписывались во многие наблюдения, принадлежавшие мужчинам

### 3.6 Нелинейные модели

В этом разделе происходит построение *нелинейных моделей* для предсказания повторных максимумов, проверяются формулы МакГлотина, Ломбарди 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  $RM = MRM \cdot (1 + c \cdot Count)$  с разными значениями коэффициента  $c$ . Как выяснилось, эти отличия обусловлены только тем, на каком разбросе повторений происходил подбор  $c$ : 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 (Рис. 62).

Используемый диапазон повторений	Оценка параметра
не более чем 2-3	0.0353338
не более чем 4-6	0.0314487
не более чем 7-10	0.0299178
не более чем 11-20	0.0259797
любой	0.0247547

Рисунок 62

Обучив такую модель (с названием  $n_1$ ) для диапазона повторений 2-10, получим результаты в 30% ошибок.

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

$$RM = MRM \cdot \frac{a}{b - Count} \quad (8)$$

с разными  $a, b$ . Обучив такую модель (с названием  $n_2$ ) для диапазона повторений 2-10, получим более точную модель (20% ошибок).

Аналогично модели Мэйхью и Ватана — это вариации модели

$$RM = \frac{100 \cdot MRM}{a + b \cdot e^{-c \cdot Count}} \quad (9)$$

Обучив такую модель ( $n_3$ ), получим примерно такие же результаты, как у  $n_2$ .

Наконец, модель Ломбарди имеет вид

$$RM = MRM \cdot Count^a \quad (10)$$

Обучив её ( $n_4$ ), получим примерно те же результаты (с параметром, очень близким к 0.1, предложенным Ломбарди):

### 3.6.2 Новые предложения

Кроме этого, я попробую модель  $n_5$ :

$$RM = MRM^a \cdot \frac{b}{c - Count + d \cdot Count^2} \quad (11)$$

и  $n_6$ :

$$RM = MRM^a \cdot (b + c \cdot Count^d), \quad (12)$$

а также версии предыдущих шести моделей с поправкой коэффициентов на разные факторные переменные и с включением/отсутствием слагаемого  $\left(\frac{MRM}{Index}\right)^6$  (для линейных моделей оно играло роль).

### 3.6.3 Сравнение нелинейных моделей

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

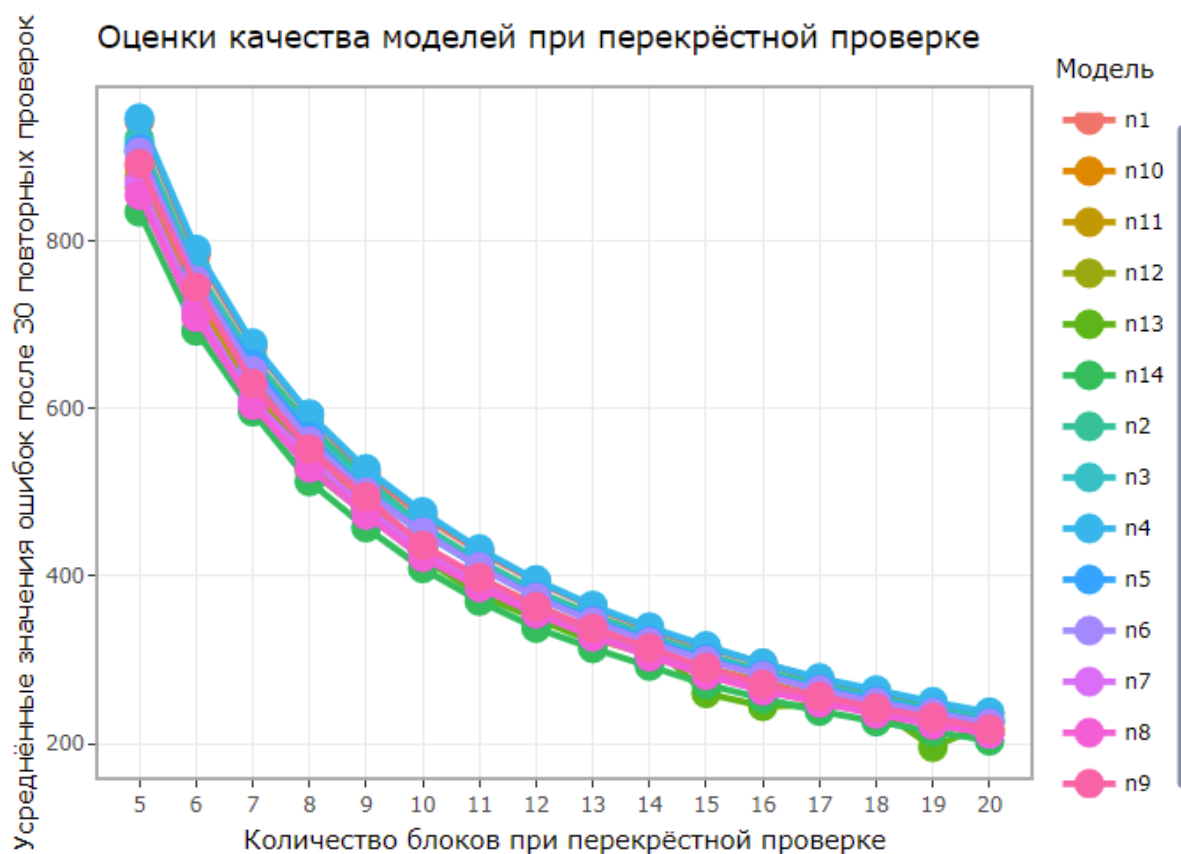


Рисунок 63

Как видно, самой лучшей моделью оказалась  $n_{14}$  (основанная на моделях Мэйхью и Ватана), на втором месте идёт  $n_8$  (моё предложение):

Обе модели имеют одинаковую среднюю ошибку в менее чем 2.9%. По точности они схожи с линейными моделями  $b_3, b_5$ .

Если сравнить лучшие найденные линейные и нелинейные модели, окажется, что модели  $n_8$  и  $b_3$  в целом неотличимы, а модель  $n_{14}$  немного превосходит  $b_5$  (Рис. 64), но не факт, что значимо.

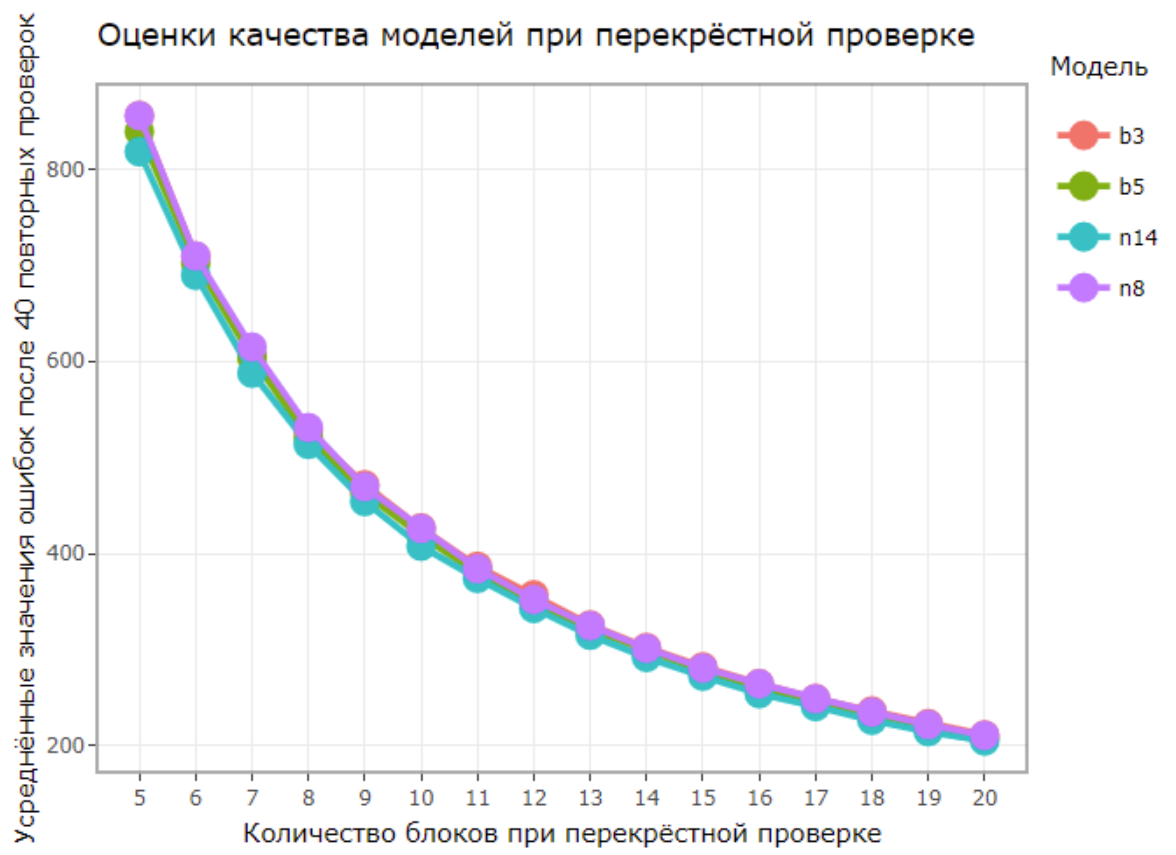


Рисунок 64

## **ЗАКЛЮЧЕНИЕ**

В результате проведённого исследования<sup>2</sup> была<sup>2</sup> 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 <https://dmitrypasko.shinyapps.io/RMbyMRMestimating/>.

Версия этой работы со многими подробностями программной реализации, интерактивными графиками и более расширенной статистикой доступна по ссылке [https://rpubs.com/DemetryPascal/RMbyMRMrus\\_2](https://rpubs.com/DemetryPascal/RMbyMRMrus_2).<sup>1</sup>

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Хэтфилд Ф. К., Всестороннее руководство по развитию силы // Красноярск: Союзспорт, 1992. - 284 с.
- 2 Вадим Протасенко. Думай! Или «Супертренинг» без заблуждений.
- 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 Хасты 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 «The Cube method» (2012)
- 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. <sup>1</sup>