**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ**

**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»**

**(РУТ (МИИТ)**

Институт/факультет «Академия “Высшая Инженерная Школа”, АВИШ»

Специальность/Направление подготовки Информатика и вычислительная техника

**КУРСОВАЯ РАБОТА**

**по дисциплине:** Технологии хранения больших данных

|  |  |
| --- | --- |
| **на тему:** | "Аналитика успеваемости студентов с помощью методов Big Data " |

|  |  |
| --- | --- |
| **Студент группы ШАД 21**2 | **(Гаганов А. А. )** |
| **Научный руководитель** | **(Зуев И. А. )** |

**Москва 2023 г.**

# **Оглавление**

[**Оглавление** 2](#_Toc153196240)

[**Введение** 3](#_Toc153196241)

[**Исследование графиков построенных в ходе работы** 4](#_Toc153196242)

# **Введение**

Тема успеваемости студентов всегда будет актуальной, ввиду её большого влияния на будущее как самих студентов, так и образовательных учреждений. Целью данной курсовой работы является проведение аналитического исследования данных об успеваемости студентов с целью выявления основных факторов, влияющих на неё.

В работе будет осуществлён анализ различных аспектов, влияющих на успеваемость, включая поддержку в учёбе со стороны семьи, здоровье студента и даже наличие у него романтических отношений, а также многое другое.

Результаты исследования позволят выяснить ключевые факторы, влияющие на успеваемость студентов, а также предложить способы по её улучшению. Полученные выводы могут быть полезны для преподавателей, самих студентов, а также их родителей, помогая им понять, что именно влияет на оценки, и решить, как их можно улучшить.

В ходе работы будут использованы соответствующие теоретические и практические материалы, чтобы достичь поставленных целей.

# **Исследование графиков построенных в ходе работы**

Для получения описательной статистики был взят датасет, в котором было приведено исследование, по итогу которого были описаны переменные и проверены данные на нормальность. После начался этап построения графиков. Опишем эти графики:

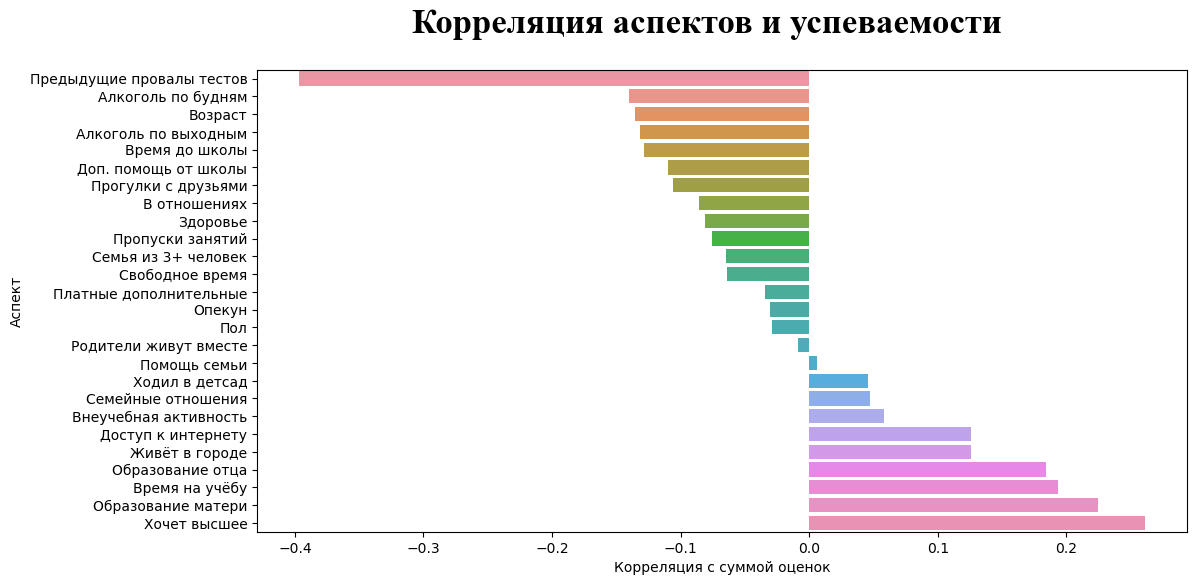


Рис. (1)

Вывод: в используемом датасете можно выделить 26 аспектов, влияющих на успеваемость учеников. Наибольшее влияние имеют количество предыдущих провалов тестов, желание получит высшее образование, время, выделенное на учёбу в неделю, и образование родителей.

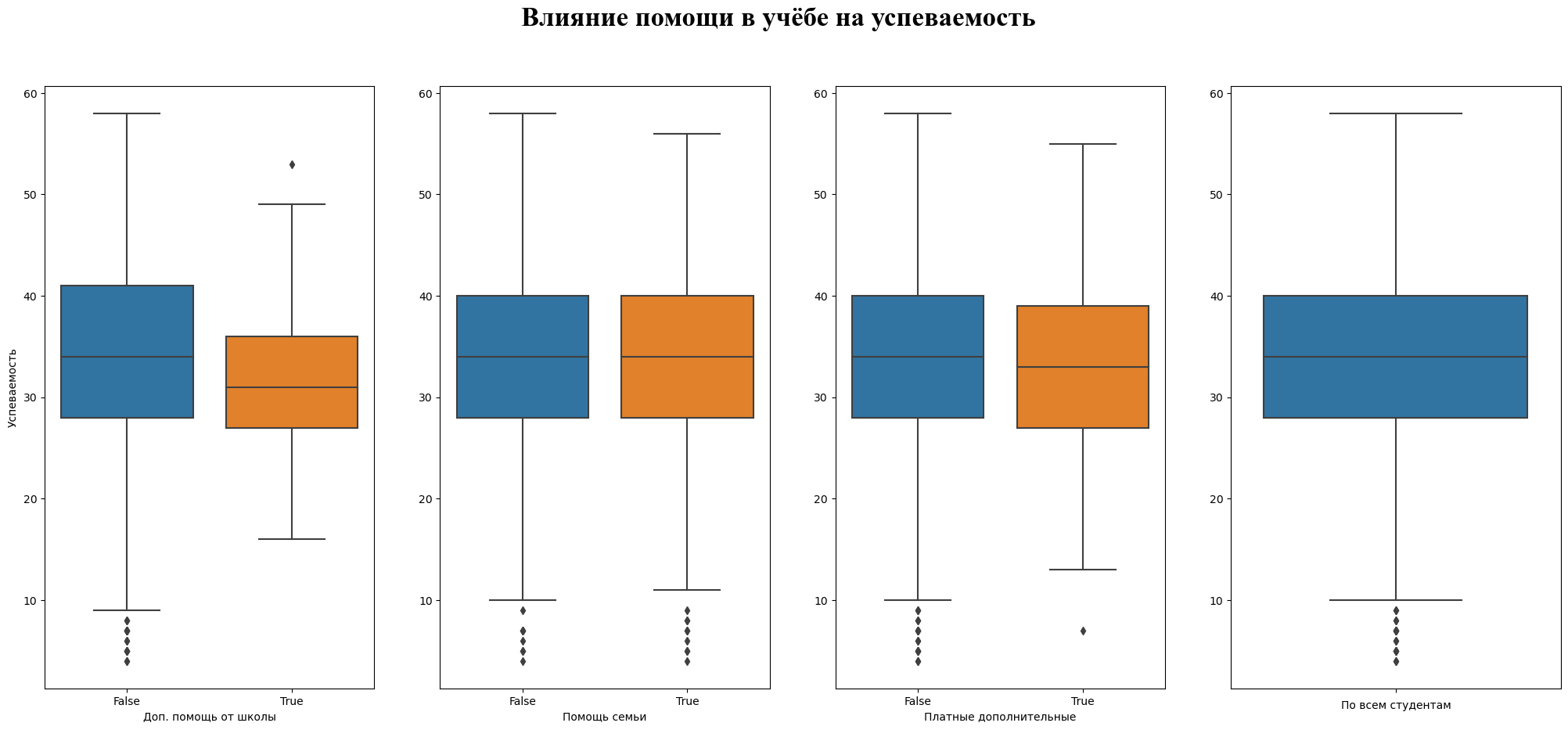


Рис. (2)

Вывод: помощь в обучении действительно может иметь положительный эффект: студенты, получающие помощь имеют более высокий порог минимальной суммы балов, но также стоит заметить, что лучшие студенты находятся именно среди тех, кто помощь не получает. Отсюда делаем логичный вывод: помощи «требуют» и, соответственно, получают лишь «отстающие» студенты, а те, кто учится действительно хорошо, справляются и без неё.

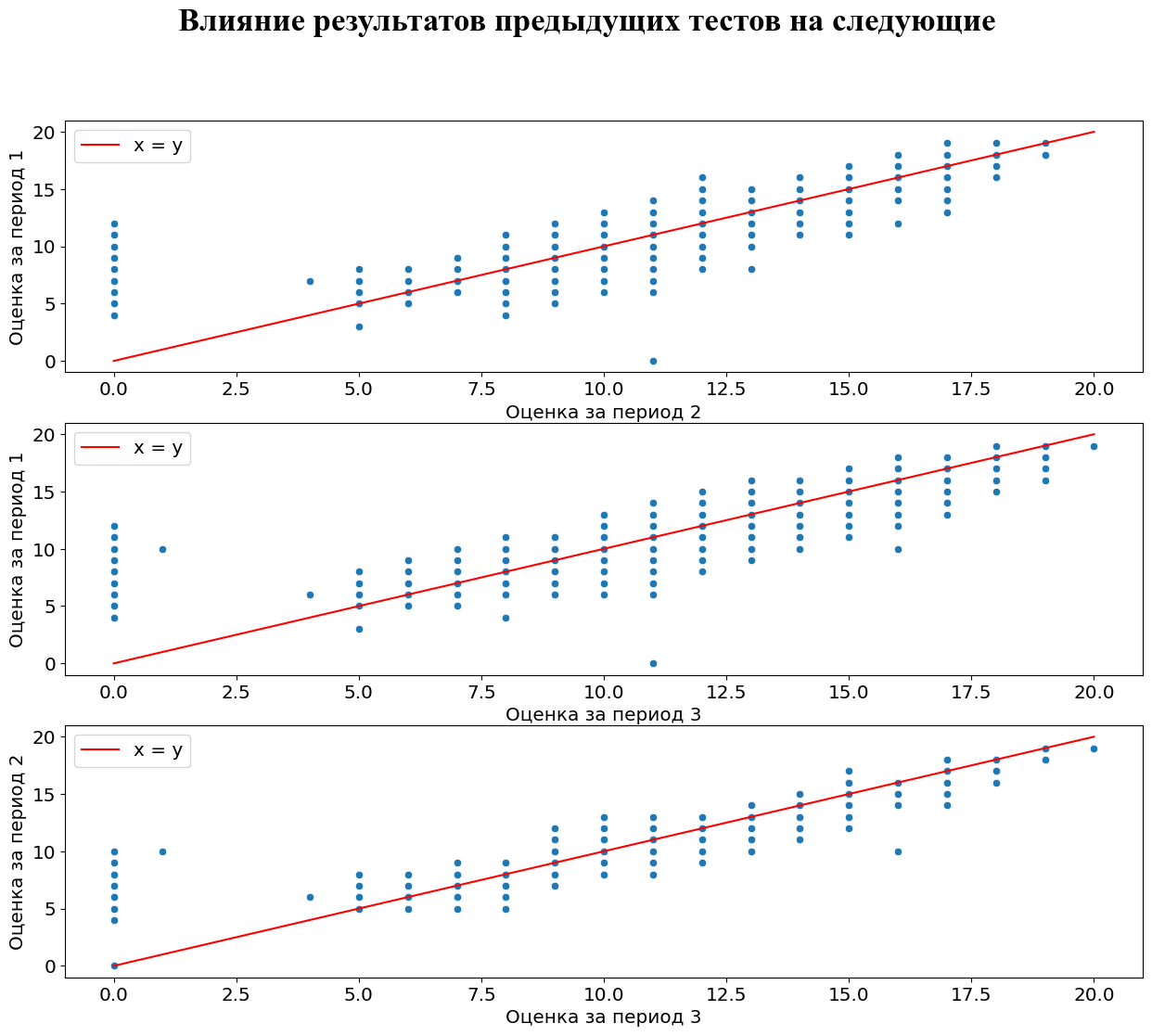


Рис. (3)

**Вывод**: не имеет значения, какие 2 теста взять, результат один – как студент сдал предыдущий тест – примерно так же он сдаст и следующий. Нельзя сказать, что кто-то «решил взяться за ум» после провала на прошлом тесте (в таком случае большинство точек было бы ниже красной линии), так же, как и нельзя говорить об обратном, что ученики «скатываются» в плохие оценки (в таком случае большинство точек было бы выше красной линии).

Наблюдение: присутствуют так же выбросы, представленные студентами, которые сдавали предыдущий тест и по какой-либо причине не сдавали следующий, причём обратной ситуации практически не наблюдается. Вероятно так выглядят студенты, покинувшие школу после теста, так как среди тех, кто отсутствовал на 2-м тесте нет ни 1, кто присутствовал бы на 3-м, а тех, кто получил 0 за 3-й стало больше, чем тех, кто не получил балов за 2 тест, вероятно, потому что они покинули школу позже первой группы. Ну и то, что балы за первый тест ни у кого из покинувших школу нельзя назвать выдающимися, подкрепляет эту теорию (возможно, они вообще были отчислены).

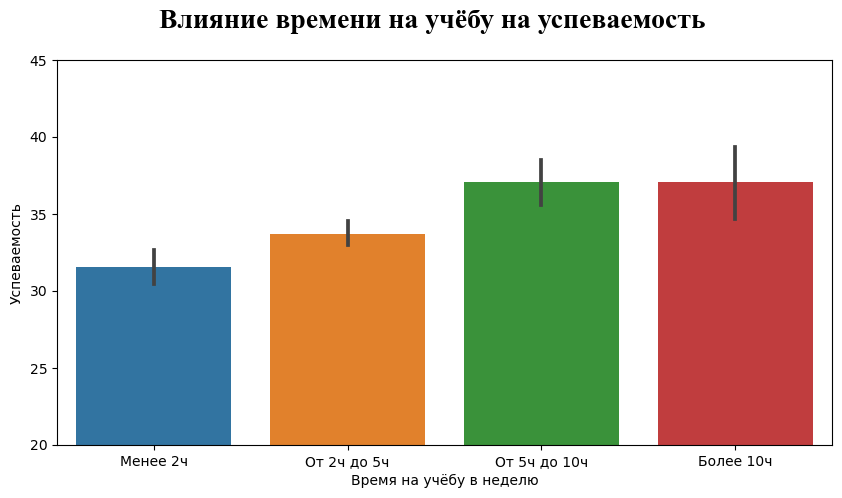


Рис. (4)

**Вывод:** оптимальное время на учёбу лежит в районе от 5 до 10 часов, потому что средний бал практически не меняется, а прирост максимального становится несущественным, а что самое главное – минимальный бал заметно падает, что, вероятно, является следствием «выгорания» студентов, переутомлённых учёбой, а так же указывает на то, что зубрёжка по несколько часов в день не сделает студента умнее.

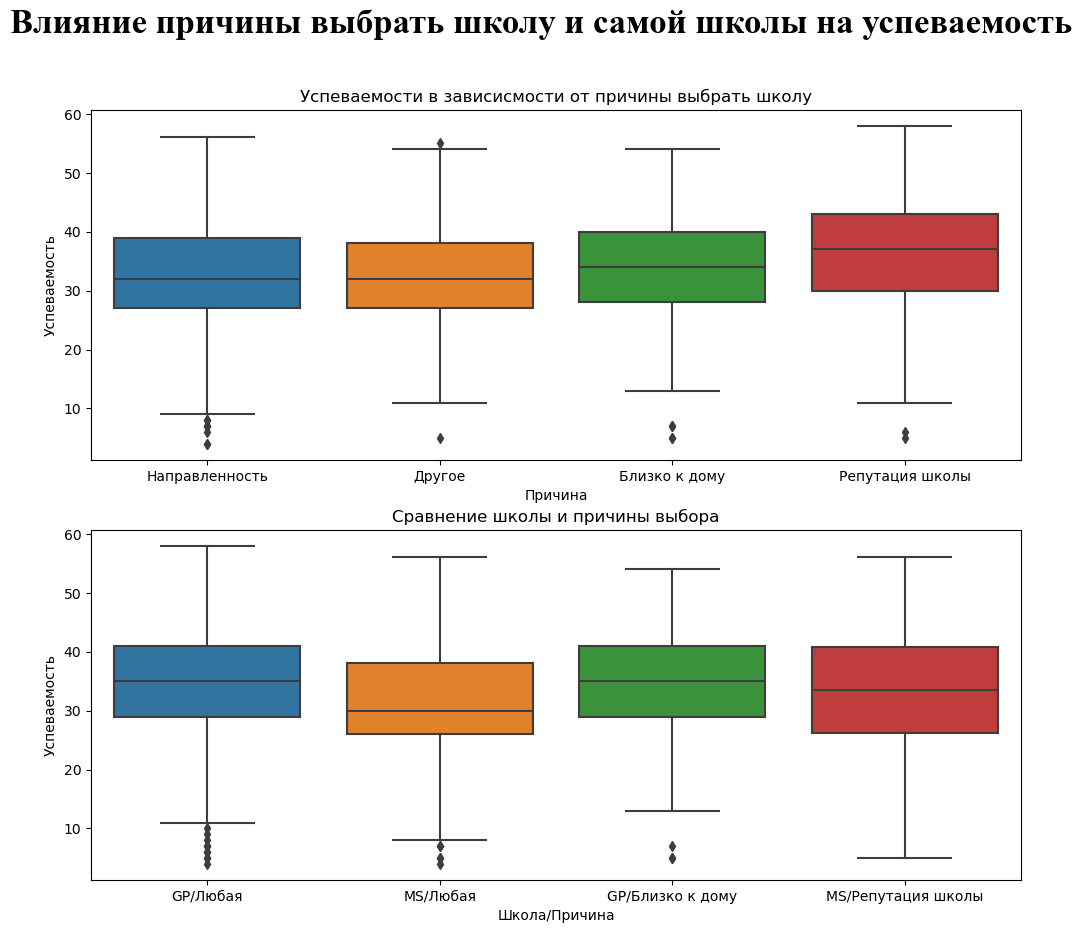


Рис. (5)

**Вывод:** причина выбрать школу имеет немалый эффект на успеваемость, вероятно, из-за её прямой связи с мотивацией на учёбу, пример: выбрал школу из-за её репутации – стараешься соответствовать – учишься лучше. Эффект достаточно сильный, чтобы практически нивелировать разницу в силе преподавания в разных школах. Хоть школа GP и сильнее в целом, если сравнить её учеников с условно низкой мотивацией и учеников школы MS с условно высокой – то разница станет сильно меньше, а максимум успеваемости и того станет выше.

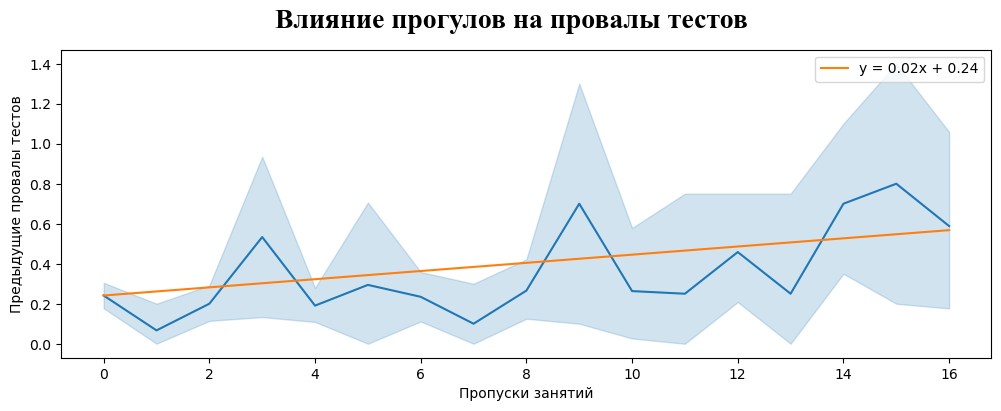


Рис. (6)

**Вывод:** пропуск занятий имеет влияние на провалы тестов, а следовательно, и успеваемость. Хоть рост среднего провалов и не велик, но разброс становится сильно больше, а следовательно пропуски помимо очевидного ухудшения успеваемости влекут так же и её нестабильность.

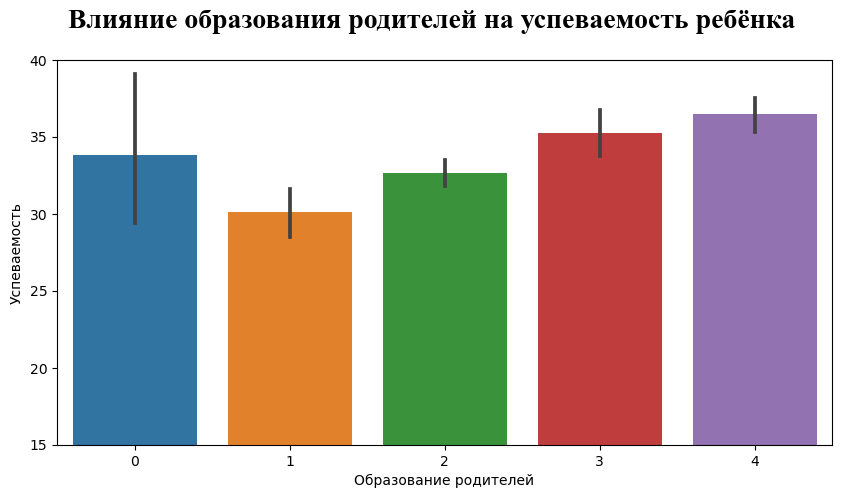
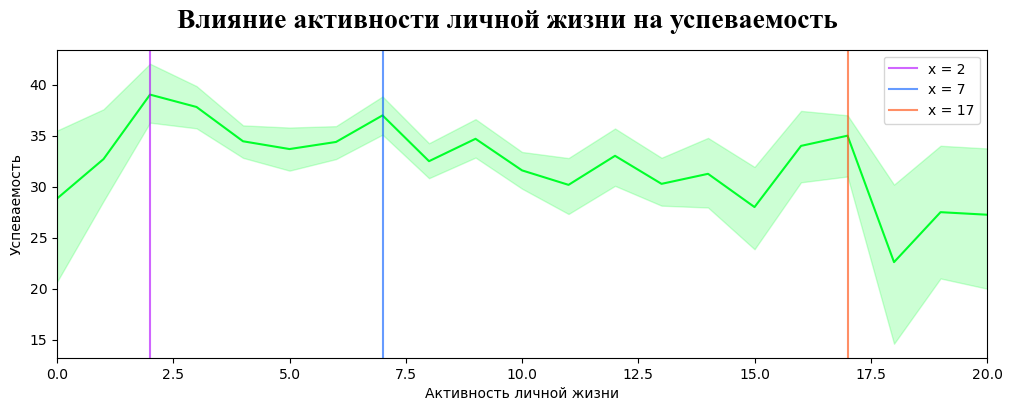


Рис. (7)

**Вывод:** хоть и, смотря на рис. (1), можно было сделать вывод, что образование родителей имеет достаточное положительное влияние на успеваемость их ребёнка, на деле можно увидеть, что тенденция не такая однозначная. Если смотреть на значения с 1 по 4, то действительно видно, что чем лучше образование родителей – тем и выше успеваемость ребёнка, но отсутствие образования у обоих родителей показывает показатели почти на уровне 3, и это даже не смотря на максимальную успеваемость, которая у детей с необразованными родителями самая высокая. Возможно, это связано с тенденцией родителей обеспечить своим детям «жизнь лучше, чем была у них», из-за чего люди совсем без образования стараются обеспечить его для своих детей. Так же, возможно, условия жизни, которые в таком случае не самые благоприятные, поднимают мотивацию детей к учёбе.



**Вывод:** максимум успеваемости достигается при активности равной 2 из 20,