



Федеральное государственное автономное образовательное учреждение высшего  
образования

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ  
ШКОЛА ЭКОНОМИКИ»

Факультет экономических наук, образовательная программа «Экономика»

Домашняя работа по курсу «Эконометрика – 2»

*“Исследование влияния объективных факторов на потребление алкоголя в средней  
школе”*

Выполнили:

Сокуров Рустам Астемирович

Родионов Илья Владимирович

Грицкова Марина Вадимовна

# Введение

## Актуальность

Потребление алкоголя среди подростков остается серьезной проблемой, привлекающей значительное внимание исследователей и специалистов. Не только объемы употребления алкоголя вызывают обеспокоенность: согласно данным из 2021 года, каждый пятнадцатилетний подросток в среднем потребляет около 6 литров чистого алкоголя в год по всему миру. Однако, принимая во внимание, что 61,7% населения в возрасте 15 лет воздерживаются от употребления алкоголя в течение последних 12 месяцев, этот показатель увеличивается до 14,63 литров в год [*Topics in Public Health, 2021*]. Важно отметить не только количественные аспекты, но и момент начала употребления: около 20% подростков в Соединенных Штатах начинают пить к 13 годам, согласно отчету Центра по контролю и профилактике заболеваний (CDC).

Потребление алкоголя среди подростков негативно сказывается не только на их здоровье, но и приводит к серьезным последствиям, таким как автомобильные аварии [*Reducing underage drinking, 2004*], злоупотребление наркотиками [*The surgeon general's call to action to prevent and reduce underage drinking, 2007*], создание риска для возникновения опасных сексуальных связей [*Binge drinking and associated health risk behaviors among high school students, 2007*], а также негативно влияет на успеваемость в школе и аттестационные оценки [*High school classmates and college success. Sociology of Education, 2009*].

Существует обширное количество исследований, посвященных факторам, влияющим на алкогольную зависимость в подростковом возрасте. В большинстве своем эти исследования сфокусированы на социальных факторах, оказывающих влияние на уровень потребления алкоголя среди подростков. Среди них находятся исследования, рассматривающие влияние отношений с ровесниками и стабильность отношений с родителями, а также другие социальные аспекты [*Social network analysis: methods and applications, 1994; Close friend and group influence on adolescent cigarette smoking and alcohol use, 1997; Chain reactions in adolescents' cigarette, alcohol and drug use, 2004*]. Кроме того, были разработаны модели, уделяющие внимание взаимосвязи этих факторов с уровнем потребления алкоголя [*Introduction to stochastic actor-based models for network dynamics, 2010; The SAGE handbook of social network analysis, 2011*].

Целью данного исследования является разработка модели линейной регрессии, способной предсказывать уровень потребления алкоголя на основе преимущественно объективных факторов, которые могут быть измерены независимо от желания

ребенка сотрудничать и без необходимости глубокого погружения в его социальную среду. Это исследование может предоставить ценные данные для дальнейших исследований, поскольку изучение объективных признаков проще по сравнению с социальными факторами. Результаты данного исследования также могут быть полезны родителям, беспокоящимся о своих детях. Однако в первую очередь необходимо определить, возможно ли прогнозировать уровень употребления алкоголя на основе объективных факторов и с какой степенью уверенности это возможно сделать.

В данном исследовании авторы поставили перед собой следующие задачи:

1. Сбор актуальной научной литературы по теме исследования
2. Формирование списка объясняющих переменных, которые в дальнейшем будут включены в модель, на основе имеющейся литературы
3. Предварительный анализ имеющихся данных
4. Составление модели линейной регрессии и последующая её оценка
5. Формулирование выводов, полученных в ходе исследования

## Данные и методы

Данные для данного исследования были собраны в течение 2005-2006 учебных годов в двух государственных школах Алентежу в Португалии. Информация собиралась из школьных отчетов, а также анкет, используемых для получения дополнительной информации, которые включали в себя закрытые вопросы об образовании родителей, семейном доходе и др. Были опрошены 788 учащихся, позже 111 ответов были отброшены из-за отсутствия идентификационных данных. В результате удалось собрать два набора данных, один из которых связан с уроками по математике и содержит 395 наблюдений, а второй - с уроками португальского языка и имеет 649 наблюдений. Собранная информация использовалась для дальнейшего анализа и построения модели в работе «*Using Data Mining To Predict Secondary School Student Performance*» за авторством исследователей из Университета Минхо. В данном исследовании мы собираемся провести Тест Чоу и по его результатам определить, стоит ли остановиться на одной подборке или объединить данные по математике и испанскому языку, построив совместную модель.

Далее мы приведём описание всех переменных, что были представлены в изначальном датасете с сайта Kaggle. При выборе переменных мы руководствовались логикой и найденными научными статьями. В статье [*Adolescent alcohol use and parental and adolescent socioeconomic position in six European cities*] авторы обращают внимание на образование и благосостояние родителей респондентов, поэтому в нашем исследовании мы также рассматриваем уровень образования родителей студента, а также семейный статус родителей (проживают они вместе или раздельно)

в качестве объясняющих переменных. В той же статье описываются такие факторы, оказывающие влияние на потребление алкоголя подростками, которые отвечают за успехи студентов в учёбе. Поэтому мы решили оставить в данных такие переменные, как время, проведённое за учёбой, наличие дополнительных внешкольных занятий, количество пропущенных занятий, а также количество прошлых учебных неудач. Эти переменные могут в целом достоверно отразить отношение к учёбе определённого студента, а это в свою очередь является важным фактором при потреблении алкоголя согласно исследованию. Также в данной статье есть базовые переменные, такие как возраст и пол респондентов. В статье «Factors Influencing Adolescent Alcohol Consumption: Parents And Depression» авторы выделяют отношения ребёнка с родителями как важный фактор, влияющий на психологическое состояние подростка, что в свою очередь тесно связано с количеством потребляемого им алкоголя. Потому мы включили в наши данные такую переменную, как качество отношений в семье.

Нами было принято исключить некоторые переменные, которые, по-нашему мнению, не должны сильно влиять на нашу зависимую переменную. Первая переменная, которую мы исключим – это school. Как нам кажется, она не будет весомо влиять в силу того, что это всего лишь две школы в Португалии, и для проверки наших гипотез они никак нам не помогут. Следующими переменными, которые мы исключим, будут G1, G2 и G3 - они показывают оценки по математике или португальскому за первый и второй семестр и итоговую оценку соответственно. По-нашему мнению, у нас и так есть переменные, которые более точно показывают отношение респондентов к учёбе. Также мы решили исключить переменные Mjob и Fjob, которые показывают какая работа у родителей, потому что мы посчитали что значения этих переменных не несут какую-то весомую информацию.

Оставшиеся переменные, не описанные в найденной нами научной литературе, мы решили не исключать из модели, так как предположили, что они имеют определённое влияние на зависимую переменную, но не знали, какое именно.

#### Переменные:

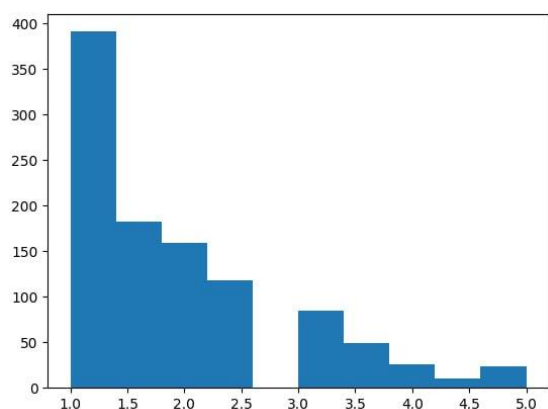
1. Sex – Пол студента (F – female; M – male). В нашем исследовании мы переведём переменную в вид 1 и 0 и возьмём её как дамми.
2. Age – Возраст студента. Предположительно, чем выше возраст студента, тем больше он пьёт.
3. Address – место, где живёт студент, в городе или в сельской местности (U – город; R – сельская местность). Эту переменную мы также возьмём как дамми. Предположительно, если студент из сельской местности, то он пьёт больше, чем тот, кто живёт в городе.

4. Famsize – размер семьи (бинарная переменная, LE3 – в семье 3 или меньше людей; GT3 – в семье больше 3 людей). В случае данной переменной у нас нет однозначных предположений о её влиянии на зависимую переменную
5. Pstatus – Бинарная переменная показывающая живут ли родители вместе (Т – живут вместе; А – живут раздельно) Также возьмём переменную как Дамми. Предположительно, если родители живут раздельно, то студент пьёт больше чем в ином случае
6. Medu – образование матери (0 – нет; 1 – начальное образование; 2 – 5–9 классов; 3 – среднее образование; 4 – высшее образование). Здесь мы предполагаем что образование каждого родителя влияет негативно на потребление алкоголя (чем оно выше)
7. Fedu – образование отца (0 – нет; 1 – начальное образование; 2 – 5–9 классов; 3 – среднее образование; 4 – высшее образование)
8. reason – причина выбора школы (home – близко к дому; reputation – репутация школы; course – учебные предпочтения; other – другие причины) В случае данной переменной у нас нет однозначного предположения о её влиянии на зависимую переменную
9. guardian – опекун студента (mother – мать; father – отец; other – другие) В случае данной переменной у нас нет однозначного предположения о её влиянии на зависимую переменную
10. traveltime – время пути до школы (1 – <15 минут; 2 – от 15 до 30 минут; 3 – от 30 минут до 1 часа; 4 – >1 часа) В случае данной переменной у нас нет однозначного предположения о её влиянии на зависимую переменную
11. studytime – время которое ученик учится в неделю, помимо школы, нам кажется здесь имеется ввиду время затраченное на занятия домашней работой и т д (1 – <2 часов; 2 – от 2 до 5 часов; 3 – от 5 до 10 часов; 4 – >10 часов). Мы предполагаем, что время затраченное на учёбу негативно влияет на потребление алкоголя.
12. failures – число провалов/неудач в прошлых классах, мы предположили, что в этом случае речь идёт о том что студенты получили неудовлетворительные оценки и имели необходимость закрывать академ задолженность ( 1–3 – количество неудач от 1 до 3; 4 – если 4 и более). Мы предполагаем, что количество неудач положительно влияет на потребление алкоголя.
13. schoolsup – поддержка, которую студентам оказывают в школе (бинарная: да или нет). Будет преобразована в дамми. Говоря о поддержке в обоих случаях что в школе, что в семье, мы предполагаем, что её наличие негативно влияет на потребление алкоголя.
14. famsup – поддержка, которую студентам оказывают в семье (бинарная: да или нет). Будет преобразована в дамми.

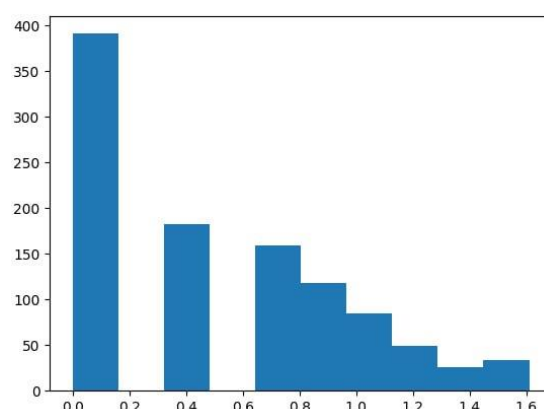
- 15.paid – наличие дополнительных платных курсов по математике или португальскому, бинарная переменная, будет преобразована в Дамми. Мы предполагаем, что наличие платных курсов негативно влияет на потребление алкоголя
- 16.activities – дополнительные внеклассные занятия, бинарная переменная, будет преобразована в Дамми. Мы считаем, что наличие внеклассных занятий влияет негативно на потребление алкоголя.
- 17.nursery – посещал ли ребёнок детский сад, бинарная переменная. Сделаем из неё Дамми. В случае данной переменной у нас нет однозначного предположения о её влиянии на зависимую переменную
- 18.higher – хочет ли студент получать высшее образование, бинарная переменная, которую мы преобразуем в Дамми. Мы предполагаем, что наличие желания получить высшее образование негативно влияет на потребления алкоголя.
- 19.internet – есть ли дома доступ к интернету, бинарная переменная, которую мы преобразуем в Дамми. Мы предполагаем, что наличие интернета негативно влияет на потребление алкоголя.
- 20.romantic – состоит ли студент в отношениях, бинарная переменная, которую мы преобразуем в Дамми. В случае данной переменной у нас нет однозначного предположения о её влиянии на зависимую переменную
- 21.famrel – качество отношений в семье (от 1 до 5, где 1 – это очень плохо, а 5 – отлично). Предположительно, качество отношений влияет негативно на количество употребляемого алкоголя.
- 22.freetime – свободное время после школы (от 1 до 5, где 1 – это очень мало, а 5 – очень много). Мы предполагаем, что свободное время положительно влияет на то, сколько выпивает студент.
- 23.goout – насколько часто студенты гуляют с друзьями (1 – очень редко, 5 – очень часто). Нам кажется, что частота прогулок с друзьями положительно влияют на то сколько студент пьёт.
- 24.health – текущее состояние здоровья (1 – очень плохо; 5 – очень хорошо).
- 25.absences – количество пропусков занятий. Мы предполагаем, что количество пропусков положительно влияет на то, сколько студент пьёт

Переменные Dalc и Walc отвечают за потребление алкоголя в будние и выходные дни соответственно и измеряются от 1 до 5, где 1 – это очень мало, а 5 – очень много. Зависимой переменной мы решили сделать среднее значение потребления алкоголя (среднее арифметическое Dalc и Walc). Зависимая переменная будет называться alc\_consumption. Мы решили объединить переменные в одну, потому что предсказывать что-то одно было бы странно, по-нашему мнению, и таким образом, лучшим вариантом мы сочли вычисление среднего значения.

Также мы решили проверить распределение зависимой переменной и распределение логарифма зависимой переменной. Таким образом, наша зависимая переменная не распределена нормально в обоих случаях. Мы также думали применить тест Бокса–Кокса, однако мы поняли, что не сможем тогда нормально интерпретировать значимые коэффициенты и поэтому нами было принято решение оставить всё как было изначально.



*Рис. 1 Распределение зависимой переменной*



*Рис. 2 Распределение логарифма зависимой переменной*

Для поиска взаимосвязи между переменными мы построили матрицу облаков рассеивания [см. Рис. 5 в Приложении]. Однако по ним нельзя сказать имеют ли переменные взаимосвязь. Также на главной диагонали этой матрицы изображены гистограммы числовых переменных, которые позволяют определить их распределение. Так, у переменных `age`; `studytime`; `famrel`; `freetime`; `gout`; `health` распределение похоже на нормальное, в то время как у `traveltime`; `failures`; `absences`; `alc_consumption` – на экспоненциальное.

Для наших признаков мы решили построить корреляционную матрицу, чтобы проверить можно ли их использовать в качестве регрессоров в нашей модели. Мы построили две матрицы: одну для числовых переменных, а вторую для категориальных (которые преобразованы нами в дамми). В каждой из двух матриц мы видим, что значения корреляций признаков друг с другом далеки от единицы (по модулю), а это значит, что мы можем использовать их как регрессоры.

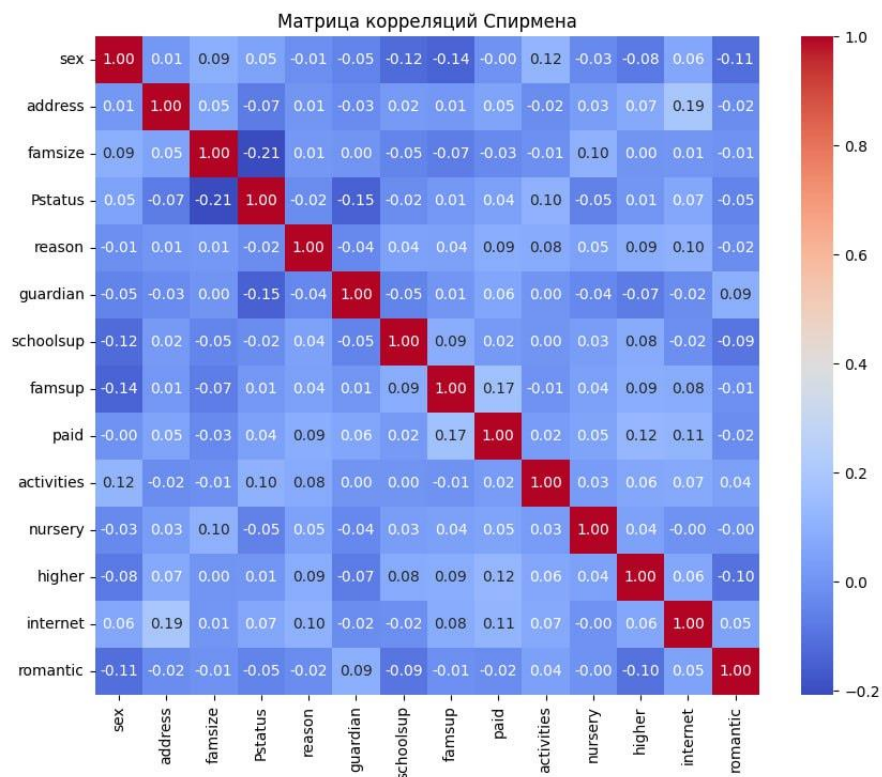


Рис. 3 Матрица корреляции для категориальных признаков

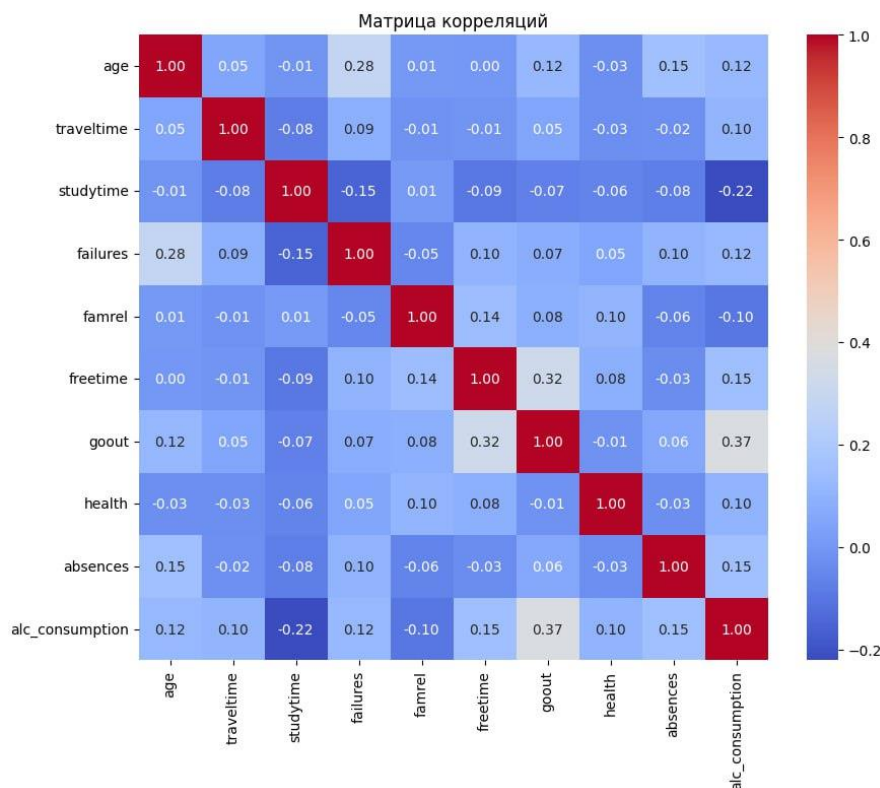


Рис. 4 Матрица корреляций для числовых признаков

Также стоит отдельно рассмотреть облака рассеивания зависимой переменной с числовыми переменными [см. Рис. 6 в Приложениях]. По ним сложно выявить какую-либо взаимосвязь зависимой переменной с независимыми числовыми переменными. Это обусловлено тем, что наша зависимая переменная не совсем непрерывная, а представляет из себя среднее двух целых чисел.



## Гипотезы

Согласно исследованию «*Alcohol consumption and binge drinking in adolescents: comparison of different migration backgrounds and rural vs. urban residence – a representative study*», количество алкоголя, потребляемого подростками, может зависеть от места их проживания, будь то сельская или городская местность. Предположительно, это может быть связано с тем, что в сельских районах существует меньше альтернативных форм проведения досуга, нежели в городах, что может способствовать более частому употреблению алкоголя среди подростков в этой местности. В связи с этим мы сформулировали следующую гипотезу: **студенты, проживающие в сельской местности, потребляют больше алкоголя, чем их сверстники из городов.** Таким образом, учитывая, что за базовое было взято значение R переменной address (студент проживает в сельской местности), коэффициент при имеющейся дамми-переменной address\_U должен иметь отрицательный знак.

В статье «Factors Influencing Adolescent Alcohol Consumption: Parents And Depression» авторы пишут, что, согласно исследованиям, качество внутрисемейных отношений оказывает значительное влияние на психологическое состояние подростка, а также на уровень стресса, который он испытывает. Так, депрессивное и тревожное состояние может быть результатом потери эмоциональной связи между родителями и ребёнком. Алкоголь же, по мнению многих людей, помогает им справиться с тревогой и депрессией и почувствовать себя лучше. Вторая наша гипотеза звучит следующим образом: **чем лучше отношения внутри семьи, тем меньше алкоголя потребляет студент.** Тогда значение коэффициента при переменной femrel должно быть отрицательным.

## Спецификация модели

В наших данных довольно много категориальных и бинарных переменных, поэтому мы решили преобразовать их в дамми-переменные. Далее мы приведём переменные, которые стали дамми и какое значение было выбрано за базовое.

- Sex – за базовое значение был выбран женский пол (female)
- Address – за базовое значение было выбрано, что студент живёт в сельской местности (R)
- Famsize – за базовое значение было взято, что в семье больше 3-х человек
- Pstatus – за базовое значение было взято, что родители живут отдельно
- Reason – базовым значением была выбрана причина “предпочтение курсов”
- guardian – базовым значением было выбрано, что опекуном ребёнка является отец.

- schoolsup – базовым значением было выбрано, что в школе не оказывают поддержку
- famsup – базовым значением было выбрано, что в семье не оказывают поддержку.
- paid – за базовое значение было взято, что у студента нет платных курсов.
- activities – базовым значением было выбрано, что у студента нет внеклассных занятий
- nursery – базовым значением является то, что студент не ходил в детский сад.
- higher – базовым значением является то, что студент не хочет получать высшее образование
- internet – базовым значением является то, что у студента нет интернета.
- romantic – базовым значением является то, что студент не состоит в романтических отношениях.

## Оценка модели

Далее будет представлена таблица ANOVA. Из-за особенностей языка Python названия переменных закодированы как их номер в модели регрессии, поэтому под таблицей будет добавлен список с расшифровкой каждой из переменных. Теперь проинтерпретируем R-squared: в нашей модели он получился 0.33. Это значит, что регрессоры, подобранные нами, объясняют 33% дисперсии зависимой переменной. Также в данной таблице есть F-stat и p-value из проверки гипотезы на адекватность модели,  $p\text{-value} = 3.91e-71$ , таким образом при любом разумном уровне значимости нулевая гипотеза отвергается, а значит наша модель адекватна.

OLS Regression Results						
Dep. Variable:	alc_consumption	R-squared:	0.330			
Model:	OLS	Adj. R-squared:	0.313			
Method:	Least Squares	F-statistic:	19.29			
Date:	Sun, 10 Dec 2023	Prob (F-statistic):	3.91e-71			
Time:	10:27:47	Log-Likelihood:	-1265.6			
No. Observations:	1044	AIC:	2585			
Df Residuals:	1017	BIC:	2719.			
Df Model:	26					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.7904	0.165	10.878	0.000	1.467	2.113
x1	0.0731	0.030	2.433	0.015	0.014	0.132
x2	0.0492	0.028	1.758	0.079	-0.006	0.104
x3	-0.0961	0.028	-3.450	0.001	-0.151	-0.041
x4	0.0212	0.028	0.746	0.456	-0.035	0.077
x5	-0.1476	0.026	-5.642	0.000	-0.199	-0.096
x6	0.0036	0.028	0.128	0.898	-0.052	0.059
x7	0.3489	0.028	12.647	0.000	0.295	0.403
x8	0.0817	0.026	3.101	0.002	0.030	0.133
x9	0.1113	0.027	4.135	0.000	0.059	0.164
x10	0.5256	0.057	9.273	0.000	0.414	0.637
x11	-0.1280	0.063	-2.035	0.042	-0.252	-0.005
x12	0.1781	0.058	3.054	0.002	0.064	0.293
x13	0.0936	0.084	1.110	0.267	-0.072	0.259
x14	0.0658	0.068	0.972	0.331	-0.067	0.199
x15	0.3000	0.090	3.332	0.001	0.123	0.477
x16	0.0068	0.069	0.099	0.921	-0.129	0.143
x17	-0.1282	0.062	-2.056	0.040	-0.251	-0.006
x18	-0.2092	0.123	-1.699	0.090	-0.451	0.032
x19	0.0155	0.084	0.186	0.853	-0.149	0.180
x20	-0.0040	0.055	-0.073	0.942	-0.112	0.104
x21	0.1722	0.065	2.639	0.008	0.044	0.300
x22	-0.0637	0.053	-1.192	0.233	-0.168	0.041
x23	-0.1996	0.065	-3.066	0.002	-0.327	-0.072
x24	-0.0145	0.100	-0.145	0.885	-0.211	0.182
x25	0.0580	0.067	0.869	0.385	-0.073	0.189
x26	0.0197	0.056	0.354	0.723	-0.090	0.129
Omnibus:	100.011	Durbin-Watson:	1.957			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	134.345			
Skew:	0.760	Prob(JB):	6.72e-30			
Kurtosis:	3.883	Cond. No.	18.4			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

## Расшифровка переменных:

1 — age	10 — sex_M	19 — schoolsup_yes
2 — traveltime	11 — address_U	20 — famsup_yes
3 — studytime	12 — famsize_LE3	21 — paid_yes
4 — failures	13 — Pstatus_T	22 — activities_yes
5 — famrel	14 — reason_home	23 — nursery_yes
6 — freetime	15 — reason_other	24 — higher_yes
7 — goout	16 — reason_reputation	25 — internet_yes
8 — health	17 — guardian_mother	26 — romantic_yes
9 — absences	18 — guardian_other	

Посмотрим на описательные статистики непрерывных переменных. Мы можем заметить, что среди них нет пропусков и нет каких-либо аномальных значений.

	age	traveltime	studytime	failures	famrel	freetime	goout	health	absences
count	1044	1044	1044	1044	1044	1044	1044	1044	1044
mean	16,726	1,523	1,970	0,264	3,936	3,201	3,156	3,543	4,435
std	1,240	0,732	0,834	0,656	0,933	1,032	1,153	1,425	6,210
min	15	1	1	0	1	1	1	1	0
25%	16	1	1	0	4	3	2	3	0
50%	17	1	2	0	4	3	3	4	2
75%	18	2	2	0	5	4	4	5	6
max	22	4	4	3	5	5	5	5	75

## Выводы

Наши гипотезы не отверглись в результате исследования на уровне значимости 5%. Таким образом, мы можем утверждать, что если студент живёт в городе, то пить он будет реже и его score потребления алкоголя будет на 0.128 ниже при прочих равных. Говоря о нашей гипотезе о влиянии отношений в семье, мы можем сделать вывод, что при прочих равных при росте score отношений в семье (то есть чем они лучше), score потребления алкоголя будет на 0.1476 меньше. Стоит отметить, что в исследованиях, которые мы нашли, исследователи получили те же результаты, что получились у нас.

## Интерпретация других значимых коэффициентов при переменных

При прочих равных условиях:

- при увеличении возраста (age) на 1 score потребления алкоголя увеличивается на 0.0731
- при увеличении времени на обучение (studytime) на 1 score потребление алкоголя уменьшается на 0.0961
- при увеличении значения, отвечающего за количество прогулок с друзьями (goout), на 1 score потребление алкоголя увеличивается на 0.3489
- при увеличении значения, отвечающего за уровень здоровья (health), на 1 score потребление алкоголя увеличивается на 0.0817
- при увеличении количества пропусков (absences) на 1 score потребление алкоголя увеличивается на 0.1113
- у мужчин score потребления алкоголя выше на 0.5256
- у студентов, у которых в семье три или меньше людей, score потребления алкоголя выше на 0.1781, чем у тех, у кого больше 3-х людей в семье
- у студентов, у которых опекуном является мать, score потребления алкоголя ниже на 0.1282, чем у тех, у которых опекуном является отец

- у студентов, которые посещают платные курсы (paid\_yes), score потребления алкоголя на 0.1722 выше, чем у тех, у кого их нет
- у студентов, которые посещали в детский сад, score потребления алкоголя на 0.1996 ниже, чем у тех, кто не посещал
- у студентов, которые не указали явно причину выбора школы, score потребления алкоголя выше на 0.3, чем у тех студентов, кто выбрал школы опираясь на предпочтительные курсы.

Хочется отметить, что в конечном итоге количество значимых коэффициентов составило примерно половину от всех переменных в составленной нами модели. Что интересно, некоторые переменные, которые нам казались важными, например, живут ли родители вместе, оказались незначимыми при любом уровне значимости. То же можно сказать и про поддержку студентов в школе и в семье: они также незначимы на любых разумных уровнях значимости. Однако нам кажется, что наша модель далеко не лучшим образом определяет количество употребляемого алкоголя. По нашему мнению, улучшить её могли бы данные о достатке семьи, о трудоустройстве родителей, о наличии иных пагубных привычек/зависимостей у респондентов, а также более развёрнутая информация о том, как учится респондент.

## Список литературы и источников

1. Topics in Public Health edited by David Claborn, 2021 - Chapter 3: Alcohol Consumption Among Adolescents by Francisca Carvajal and Jose Manuel Lerma-Cabrera
2. Centers for Disease Control and Prevention (CDC). Youth risk behavior surveillance report—United States, 2011
3. Bonnie RJ, O'Connell ME (editors). Reducing underage drinking: a collective responsibility. Washington, DC: The National Academies Press; 2004
4. U. S. Department of Health & Human Services (HHS). The surgeon general's call to action to prevent and reduce underage drinking. Rockville, MD: U.S. Department of Health and Human Services; 2007
5. Miller JW, Naimi TS, Brewer RD, Jones SE. Binge drinking and associated health risk behaviors among high school students. Pediatrics 2007
6. Fletcher JM, Tienda M. High school classmates and college success. Sociology of Education 2009
7. Wasserman S, Faust K. Social network analysis: methods and applications. New York: Cambridge University Press; 1994
8. Urberg KA, Degirmencioglu SM, Pilgram C. Close friend and group influence on adolescent cigarette smoking and alcohol use. Developmental Psychology 1997
9. Kirke DM. Chain reactions in adolescents' cigarette, alcohol and drug use: similarity through peer influence or the patterning of ties in peer networks? Social Networks 2004
10. Snijders TAB, van de Bunt GG, Steglich CE. Introduction to stochastic actor-based models for network dynamics. Social Networks 2010
11. Snijders TAB. Network dynamics. In: Scott J, Carrington PJ, editors. The SAGE handbook of social network analysis. Thousand Oaks, CA: Sage; 2011
12. Adolescent alcohol use and parental and adolescent socioeconomic position in six European cities. Marina Bosque-Prous, Mirte A. G. Kuipers, Albert Espelt, Matthias Richter, Arja Rimpelä, Julian Perelman, Bruno Federico, M. Teresa Brugal, Vincent Lorant & Anton E. Kunst 2017
13. Alcohol consumption and binge drinking in adolescents: comparison of different migration backgrounds and rural vs. urban residence - a representative study Carolin Donath, Elmar Gräbel, Dirk Baier, Christian Pfeiffer, Deniz Karagülle, Stefan Bleich & Thomas Hillemacher (2011)
14. Factors influencing adolescent alcohol consumption: parents and depression. Pavel Valedinsky, Valeriya Ivanyushina, Daniel Alexandrov, Daria Khodorenko (2023)

## Приложения

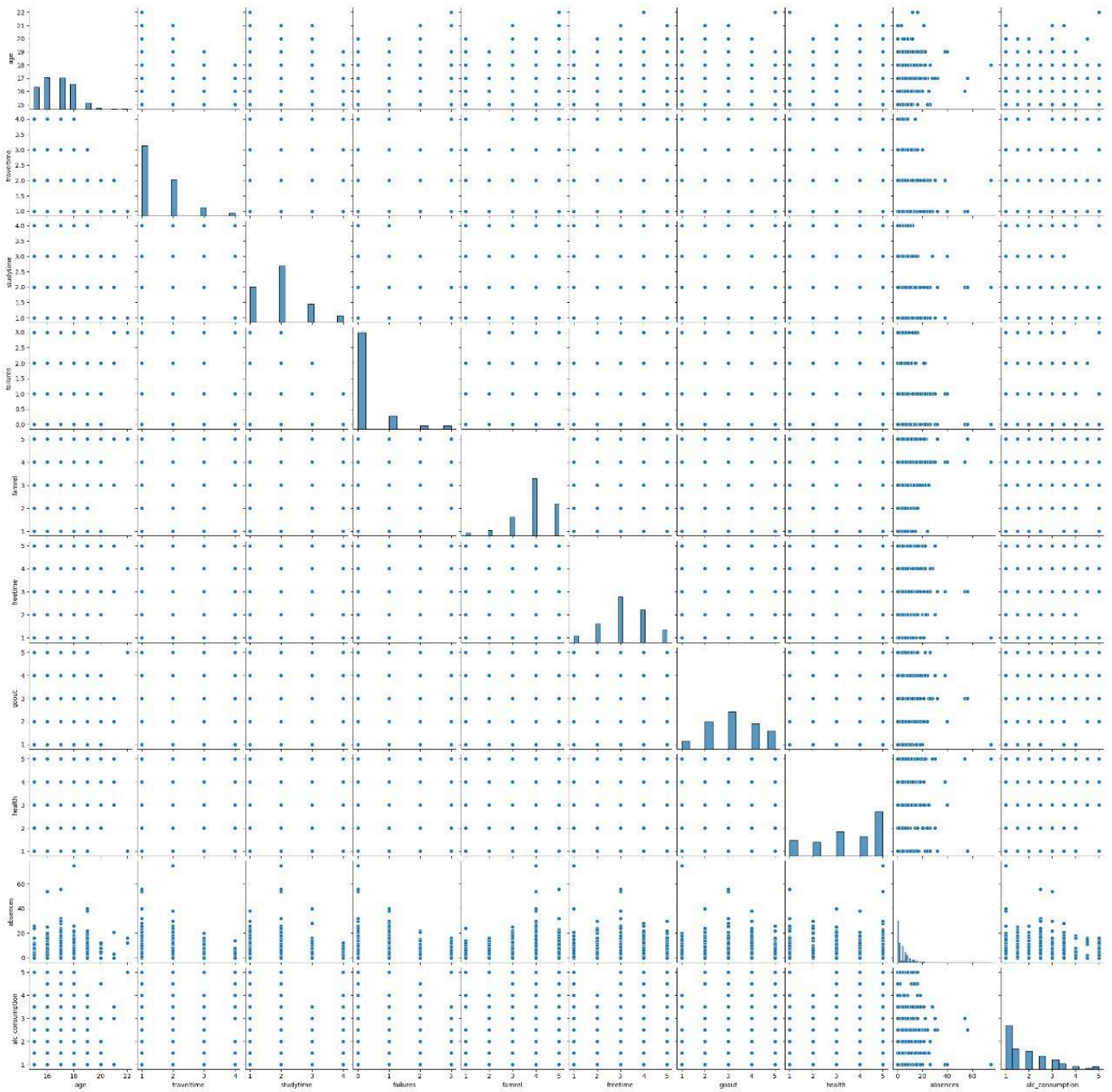


Рис. 5 Матрица облаков рассеивания

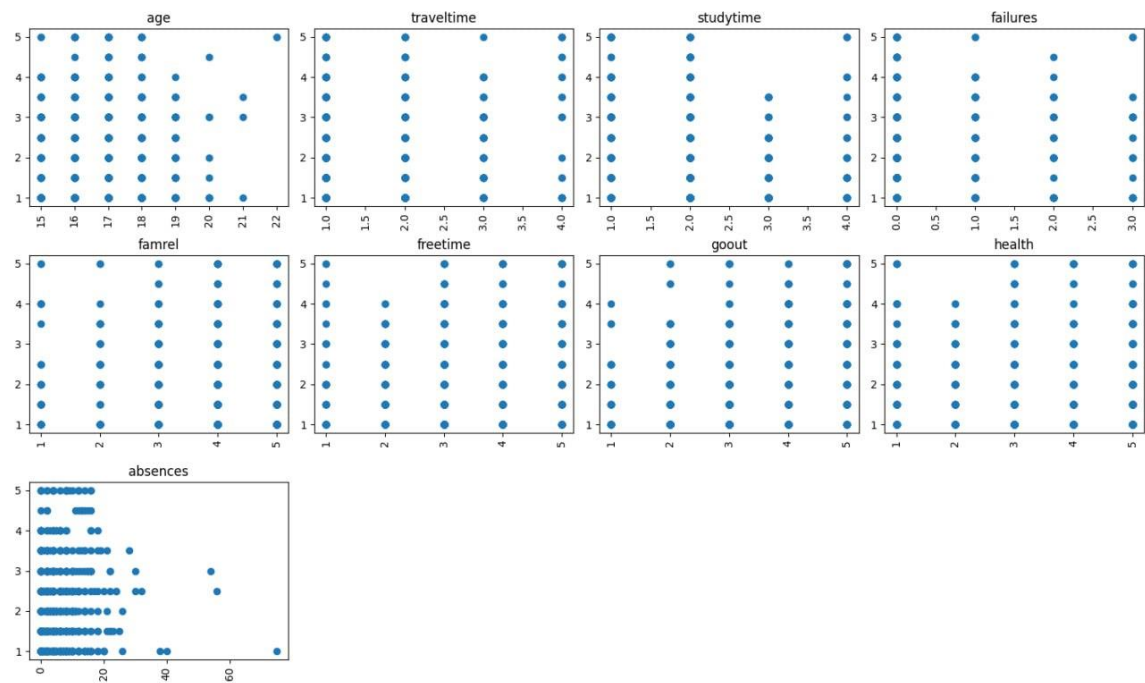


Рис. 6 Облака рассеивания для зависимой переменной и числовых признаков