Введение: цель и задачи

Рассматривается большой набор медико-психологических данных, включающий значения индекса массы тела (ИМТ), результаты психологических опросников и дополнительную информацию о респондентах.

Цель работы: применить статистические модели на основе гамма-распределения для анализа влияния факторов на параметры распределения ИМТ.

Задачи:

- Проверить согласие с гамма-распределением
- Проверить однородность параметров гамма-распределения между группами в зависимости от значений факторов.
- Расширить подход с помощью применения специальной регрессионной модели с несимметричными остатками.

Статистические модели на основе гамма распределения

└Введение: цель и задачи

Введение: цель и задачи

Рассматривается большой набор медико-психологических данных, включающий значения индекса массы тела (ИМТ), результаты психологических опросников и дополнительную информацию о респондентах.

Цель работы: применить статистические модели на основе гамма-распределения для анализа влияния факторов на параметры распределения ИМТ.

Задачи:

- Проверить согласие с гамма-распределением
- Проверить однородность параметров гамма-распределения между группами в зависимости от значений факторов.
- Расширить подход с помощью применения специальной регрессионной модели с несимметричными остатками.

В работе рассматривается крупный массив данных, включающий значения индекса массы тела, результаты психологических опросников и дополнительную информацию о респондентах.
Цель работы — применить статистические модели, основанные на гамма-распределении, чтобы изучить, как различные факторы влияют на параметры распределения ИМТ.
Такой подход особенно полезен, потому что средние значения могут совпадать между группами, но при этом сами распределения — отличаться по параметрам.
Первой задачей было проверить, насколько согласуется ли имт с гамма-распределением.
Затем — сравнить параметры гамма-распределений между группами, различающимися по уровню риска расстройств пищевого поведения.
А также расширить подход, построив регрессионную модель с гамма-распределёнными остатками.
Анализ проводится только для женщин, поскольку:
распределение ИМТ у мужчин существенно ближе к нормальному.
мужчин существенно меньше — около 1500 против 26 тысяч женщин.
данные для мужчин содержали большое количество пропусков.

Используется именно гамма распределение, так как ИМТ — это положительная величина, как правило асимметричная, с длинным правым хвостом. Гамма-распределение хорошо подходит для описания таких данных.

Общая информация о данных

Данные: 27,770 наблюдений по 38 переменным (данные получены сотрудниками ПСПбГМУ им. акад. И.П. Павлова).
Переменные:

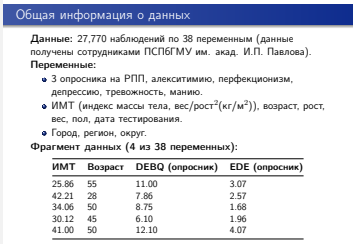
- 3 опросника на РПП, алекситимию, перфекционизм, депрессию, тревожность, манию.
- ИМТ (индекс массы тела, вес/рост^2(кг/м^2)), возраст, рост, вес, пол, дата тестирования.
- Город, регион, округ.

Фрагмент данных (4 из 38 переменных):

ИМТ	Возраст	DEBQ (опросник)	EDE (опросник)
25.86	55	11.00	3.07
42.21	28	7.86	2.57
34.06	50	8.75	1.68
30.12	45	6.10	1.96
41.00	50	12.10	4.07

Статистические модели на основе гамма распределения

└Общая информация о данных



Исходный набор данных содержит 27 тысяч наблюдений по 38 переменным — каждая строка соответствует отдельному респонденту, для которого записано:
Количество набранных баллов по различным психологическим опросникам, таким как тесты на РПП, алекситимию, депрессию, тревожность, манию и перфекционизм
Антропометрические данные: ИМТ, возраст, рост и вес, а также дополнительные сведения — пол, дату тестирования и место проживания.
На слайде показан фрагмент таблицы: данные пяти респондентов по четырём переменным — индекс массы тела, возраст и результаты двух опросников, отражающих риск расстройств пищевого поведения.
Пояснение: EDE - Eating Disorder Examination DEBQ -Dutch Eating Behaviour Questionnaire

Проверка согласия с гамма-распределением

- Рассматриваются значения ИМТ у женщин.
- Для проверки согласия распределения ИМТ с гамма, взято **20 случайных выборок объемом** $n = 200$**.**

Для каждой выборки из 20:

- Параметры гамма-распределения λ, β оценены методом максимального правдоподобия.
- Проверена гипотеза H_0: ИМТ согласуется с гамма-распределением.
- Критерий проверки: хи-квадрат Пирсона, уровень значимости $\alpha = 0.05$.

H_0 **отвергается** в 12 из 20 выборок.

Статистические модели на основе гамма распределения

└Проверка согласия с гамма-распределением

Проверка согласия с гамма-распределением

- Рассматриваются значения ИМТ у женщин.
- Для проверки согласия распределения ИМТ с гамма, взято **20 случайных выборок объемом** $n = 200$.

Для каждой выборки из 20:

- Параметры гамма-распределения λ, β оценены методом максимального правдоподобия.
- Проверена гипотеза H_0: ИМТ согласуется с гамма-распределением.
- Критерий проверки: хи-квадрат Пирсона, уровень значимости $\alpha = 0.05$.

H_0 **отвергается** в 12 из 20 выборок.

Первым шагом было проверка согласия распределения ИМТ с гамма-распределением.
Для проверки согласия было взято 20 случайных выборок по 200 человек.
Для каждой выборки параметры гамма-распределения — форма и масштаб - оценивались методом максимального правдоподобия.
Далее проверяется гипотеза H0 о согласии распределения ИМТ с гамма распределением по критерию Пирсона.
Гипотеза H0 была отвергнута в 12 из 20 выборок.

Согласие с гамма-распределением в подгруппах

Стратификация ИМТ:

- **Пищевое поведение:** 5 категорий в зависимости от значения по Голландскому опроснику пищевого поведения (DEBQ).
- **Возраст:** 4 категории (18–25, 26–40, 41–59, 60–65 лет).

Для каждой категории взято **20 выборок объемом** $n = 200$ и проверяется аналогичная гипотеза H_0, $\alpha = 0.05$.

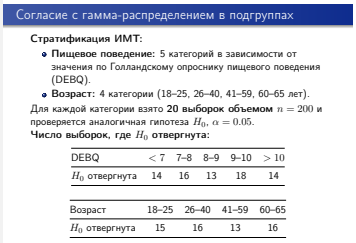
Число выборок, где H_0 отвергнута:

DEBQ	< 7	7–8	8–9	9–10	> 10
H_0 отвергнута	14	16	13	18	14

Возраст	18–25	26–40	41–59	60–65
H_0 отвергнута	15	16	13	16

Статистические модели на основе гамма распределения

└Согласие с гамма-распределением в подгруппах



Итак, на совокупной выборке обнаружилось лишь частичное согласие с гамма-распределением. Далее я разделяла данные на подгруппы — чтобы проверить, удастся ли получить лучшее согласие внутри этих подгрупп.

В первом случае, стратификация проводилась в зависимости от количества набранных баллов по Голландскому опроснику пищевого поведения (тест на РПП), данные разделены на пять категорий.

Во втором случае, стратификация — по возрасту, на четыре подгруппы.

Для каждой категории было взято 20 случайных выборок по 200 человек, и для каждой категории аналогичным образом проверялась гипотеза о согласии с гамма-распределением.

В первой таблице указаны результаты для категорий в зависимости от баллов по Голландскому опроснику пищевого поведения, во второй в зависимости от возраста.

Как можно заметить, ни в одной из категорий не удается получить согласие в большинстве выборок.

Линейная регрессия и анализ остатков

- Модель:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2),$$

где y_i — ИМТ, x_i — возраст, ε_i — остаток.

- По тесту Шапиро-Уилка p-value < 0.05 во всех выборках ($\alpha = 0.05$).

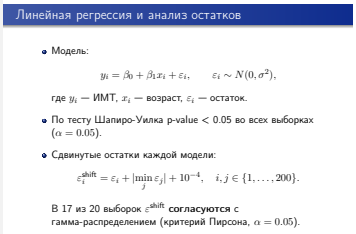
- Сдвинутые остатки каждой модели:

$$\varepsilon_i^{\text{shift}} = \varepsilon_i + |\min_j \varepsilon_j| + 10^{-4}, \quad i, j \in \{1, \ldots, 200\}.$$

В 17 из 20 выборок $\varepsilon^{\text{shift}}$ **согласуются** с гамма-распределением (критерий Пирсона, $\alpha = 0.05$).

Статистические модели на основе гамма распределения

└Линейная регрессия и анализ остатков

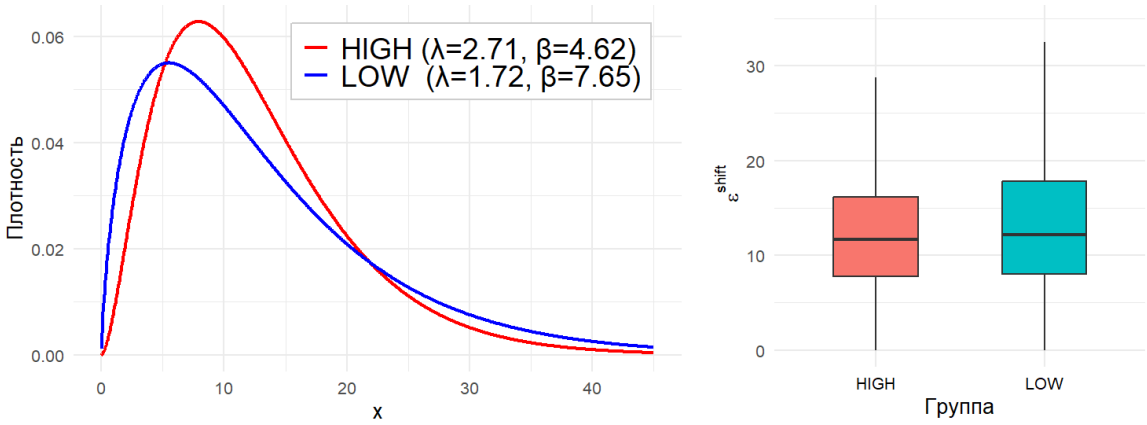


Так как даже после стратификации не удается получить устойчивое согласие с гамма-распределением, следующим шагом было построение линейной регрессионной модели.
Модель построена для каждой из 20 выборок обьемом 200, зависимая переменная - ИМТ, независимая - возраст.
Гипотеза о нормальности остатков отвергается в каждой выборке.
Далее, остатки были сдвинуты в положительную область.
И в 17 из 20 выборок сдвинутые остатки согласуются с гамма-распределением.
То есть, после устранения влияния возраста на ИМТ, можно говорить о хорошем согласии с гамма-распределением.

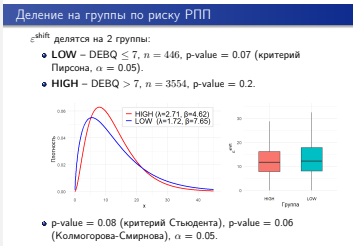
Деление на группы по риску РПП

$\varepsilon^{\text{shift}}$ делятся на 2 группы:

- **LOW** – DEBQ ≤ 7, $n = 446$, p-value = 0.07 (критерий Пирсона, $\alpha = 0.05$).
- **HIGH** – DEBQ > 7, $n = 3554$, p-value = 0.2.



- p-value = 0.08 (критерий Стьюдента), p-value = 0.06 (Колмогорова-Смирнова), $\alpha = 0.05$.



Далее сдвинутые остатки объединяются в один набор обьемом четыре тысячи и делятся на две группы в зависимости от количества набранных баллов по Голандскому опроснику пищевого поведения:
• LOW — балл $\leq$ 7, 446 наблюдений; • HIGH — балл $>$ 7, 3 554 наблюдения.
Обе группы согласуются с гамма распределением.
Из графика видно, что средние практически совпадают, но хвост распрделения в группе HIGH тяжелее. Формально среднее различие не значимо: t-тест даёт p $\approx$ 0.08.
Судя из графика плотностей, формы распределений также отличаются, однако тест Колмогорова-Смирнова — p = 0.06, тоже не значимое различие.
Можно предположить, что различия между этими группами заключаются в параметрах гамма распределения.
Поэтому далее будут проверятся однородность параметров этих групп.

Проверка однородности параметров

- **Параметр масштаба β.**
  Для $X \sim \Gamma(\lambda, \beta)$ верно $\operatorname{cov}(X, \ln X) = \beta$.
  Гипотеза H_0 проверяется по выборочным ковариациям с использованием критерия Стьюдента.
- **Параметр формы λ.**
  Выборки нормируются: $X^* = \dfrac{X}{\widehat{\beta}} \sim \Gamma(\lambda, 1)$.
  Сравниваются выборочные средние используя тот же критерий.

Результаты проверки гипотез для групп HIGH и LOW ($\alpha = 0.05$):

- $\beta_{LOW} = 7.65$, $\beta_{HIGH} = 4.62$, p-value $= 0.12$.
- $\lambda_{LOW} = 1.72$, $\lambda_{HIGH} = 2.71$, p-value $< 2.2 \cdot 10^{-16}$.

Статистические модели на основе гамма распределения

└Проверка однородности параметров

Проверка однородности параметров

- **Параметр масштаба** β.
  Для $X \sim \Gamma(\lambda, \beta)$ верно $\operatorname{cov}(X, \ln X) = \beta$.
  Гипотеза H_0 проверяется по выборочным ковариациям с использованием критерия Стьюдента.
- **Параметр формы** λ.
  Выборки нормируются: $X^* = \frac{X}{\widehat{\beta}} \sim \Gamma(\lambda, 1)$.
  Сравниваются выборочные средние используя тот же критерий.

Результаты проверки гипотез для групп HIGH и LOW ($\alpha = 0.05$):

- $\beta_{LOW} = 7.65$, $\beta_{HIGH} = 4.62$, p-value $= 0.12$.
- $\lambda_{LOW} = 1.72$, $\lambda_{HIGH} = 2.71$, p-value $< 2.2 \cdot 10^{-16}$.

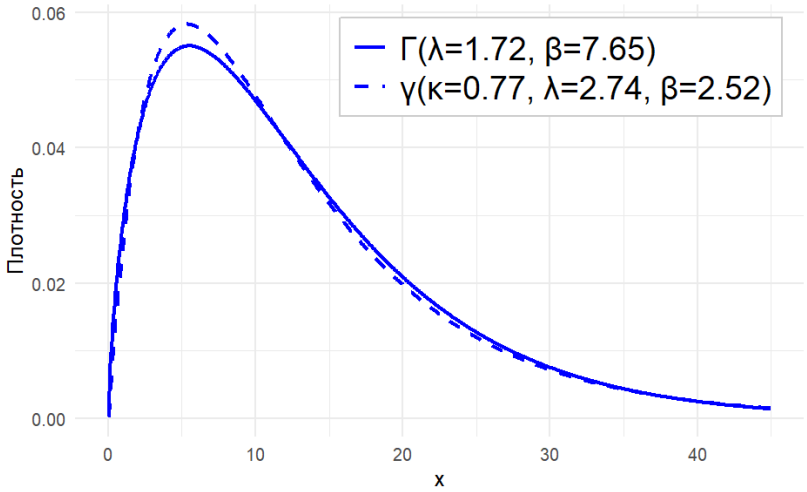
Для проверки однородности параметра масштаба бета использовалось то, что ковариация между гамма распределенной величиной и ее логарифмом равна параметру бета.
Поэтому, сравнивались выборочные ковариации этих групп с использованием критерия Стьюдента на равенство средних.
Для проверки однородности по параметру формы лямбда выборки нормировались на параметр бета. Так как для гамма распределенной величины среднее в таком случае будет равно параметру лямбда, группы сравнивались по аналогичному критерию.

В параметре масштаба значимых различий обнаружено не было, а параметр формы значимо отличается.

Степенное гамма-распределение

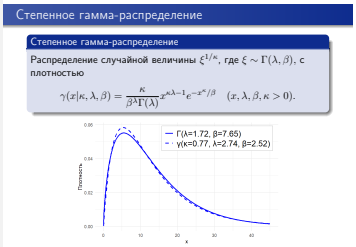
Распределение случайной величины $\xi^{1/\kappa}$, где $\xi \sim \Gamma(\lambda, \beta)$, с плотностью

$$\gamma(x|\kappa, \lambda, \beta) = \frac{\kappa}{\beta^{\lambda}\Gamma(\lambda)} x^{\kappa\lambda-1} e^{-x^{\kappa}/\beta} \quad (x, \lambda, \beta, \kappa > 0).$$



Статистические модели на основе гамма распределения

└Степенное гамма-распределение



Далее в работе использовались степенные гамма распределения, которые позволят обобщить подход и лучше охарактеризовать различия распределений.

Степенное гамма-распределение Определяется как распределение величины кси в степени один делить на каппа, где кси имеет гамма распределение.

Параметр каппа позволяет регулировать асимметрию распределения: например, при каппа<1 распределение исходной величины кси сжимается, давая более вытянутый хвост.

При каппа > 1 – наоборот, распределение менее скучено около нуля и хвост короче.

Для гамма распределения можно подобрать целое семейство степенных гамма распределений, которые будут близки к нему.

Например, на графике показана плотность распределения с параметрами, как в группе LOW, и степенное гамма распределение, параметры также указаны на графике.

Синонимичные распределения

Синонимичные распределения

Распределение P_j **синонимично** распределению P_i с заданным уровнем синонимии δ^*, если $I(i:j) < \delta^*$, где:

$$I(i:j) = H_{ij} - H_{ii}, \quad -\text{средняя информация},$$

$$H_{ij} = -\int_X f_i(x) \log f_j(x)\, dx, \quad -\text{дифференциальная энтропия}.$$

Предложение (Н. П. Алексеева, 2012 [?])

Если известны параметры степенного гамма-распределения $(\kappa_1, \beta_1, \lambda_1)$, *то, фиксируя значения* κ_2, *параметры* β_2, λ_2 *синонимичного степенного гамма-распределения находятся из системы:*

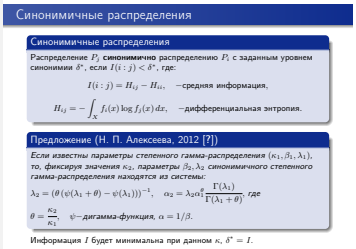
$$\lambda_2 = \left(\theta\left(\psi(\lambda_1+\theta) - \psi(\lambda_1)\right)\right)^{-1}, \quad \alpha_2 = \lambda_2 \alpha_1^{\theta} \frac{\Gamma(\lambda_1)}{\Gamma(\lambda_1+\theta)}, \text{ где}$$

$$\theta = \frac{\kappa_2}{\kappa_1}, \quad \psi - \textit{дигамма-функция}, \ \alpha = 1/\beta.$$

Информация I будет минимальна при данном κ, $\delta^* = I$.

Статистические модели на основе гамма распределения

└Синонимичные распределения



Далее, введем понятие синонимии, которое как раз используется для оценки близости распределений по информационному критерию.
На слайде определение синонимичных распределений.
В работе моего научного руководителя было доказано утверждение, которое вы можете увидеть на слайде. Оно позволяет найти параметры степенного гамма распределения при фиксированном каппа.
Причем, в таком случае информация для отличения синонимичного распределения от исходного для данного каппа будет минимальна и уровень синонимии соответственно равен и.
Пояснение: H_ij - средняя «количество информации» при кодировании значений, распределённых по закону i, с помощью модели j;
Энтропия - мера информационного содержания, неопределенности.

Номинативное распределение

При фиксированном уровне синонимии, **номинативным** называется синонимичное распределение с минимальной собственной энтропией H_{jj}.

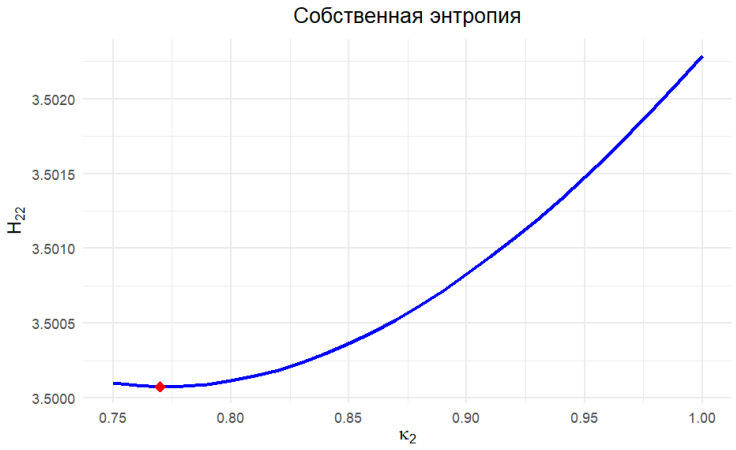


Figure: График зависимости собственной энтропии от κ для синонимичных распределений группы LOW

Статистические модели на основе гамма распределения

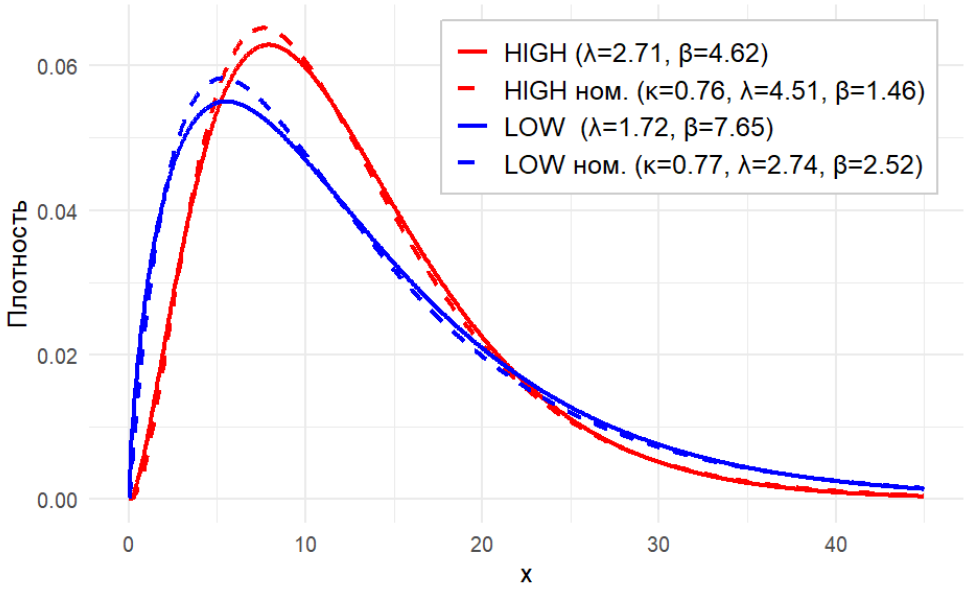
└Номинативное распределение

При фиксированном уровне синонимии, может найтись целое семейство синонимичных распределений. Распределение с минимальной собственной энтропией будем называть номинативным. На графике — зависимость собственной энтропии от параметра каппа для группы синонимичных распределений группы LOW, уровень синонимии = 0.002.

Каппа перебиралась в пределах от 0.75 до 1, так как в работе моего научного руководителя было доказано, что для гамма распределения при лямбда больше единицы, достаточно рассматривать каппа в этих пределах для нахождения минимума собственной энтропии.

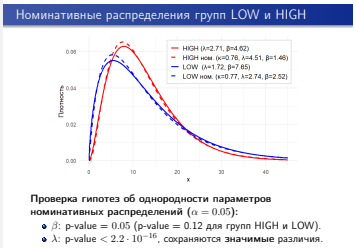
Красным показан минимум собственной энтропии синонимичного распределения - распределение с этим параметром каппа и будет номинативным.

Номинативные распределения групп LOW и HIGH



Проверка гипотез об однородности параметров номинативных распределений ($\alpha = 0.05$):

- β: p-value = 0.05 (p-value = 0.12 для групп HIGH и LOW).
- λ: p-value $< 2.2 \cdot 10^{-16}$, сохраняются **значимые** различия.



На графике показаны плотности обеих групп, плотности номинативные распределения, также указаны параметры распределений групп и параметры номинативных. Уровень синонимии - две тысячных для LOW, одна тысячная для HIGH.
Процедура проверки однородности параметров гамма распределения уже была описана ранее. Однородность параметров номинативного проверяются аналогичным образом. Различие состоит в том, что сначала данные возводятся в соответствующую степень каппа, чтобы перейти гамма распределению.
• для масштаба p-value по критерию Стьюдента равен 0.05 — совпадает с уровнем значимости, тогда как для исходных групп HIGH и LOW p-value = (0.12);
• для формы p-value < 2·10**-16. Это означает, что сохраняются значимые различия в этом параметре.
Для параметра бета: Колмогоров 0.01, Манна-Уитни 0.1.

Информационное расстояние

Симметризованное информационное расстояние между распределениями

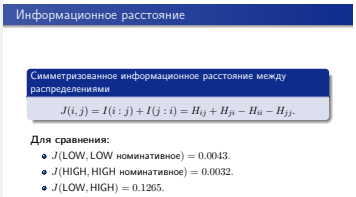
$$J(i,j) = I(i:j) + I(j:i) = H_{ij} + H_{ji} - H_{ii} - H_{jj}.$$

Для сравнения:

- $J(\text{LOW}, \text{LOW номинативное}) = 0.0043.$
- $J(\text{HIGH}, \text{HIGH номинативное}) = 0.0032.$
- $J(\text{LOW}, \text{HIGH}) = 0.1265.$

Статистические модели на основе гамма распределения

└Информационное расстояние



Чтобы описать, насколько два распределения “далеки” друг от друга, введем также симметризованное информационное расстояние, то есть сумму двух средних информаций.
Для сравнения, на слайде я привела расстояния. Распределения групп HIGH и LOW друг от друга существенно дальше, чем распределения от своих номинативных аналогов.

Figure: Информационное расстояние J между распределением и его номинативным в зависимости от объема выборки. Параметры распределения - параметры группы LOW ($\lambda = 1.72$, $\beta = 7.65$).

Статистические модели на основе гамма распределения

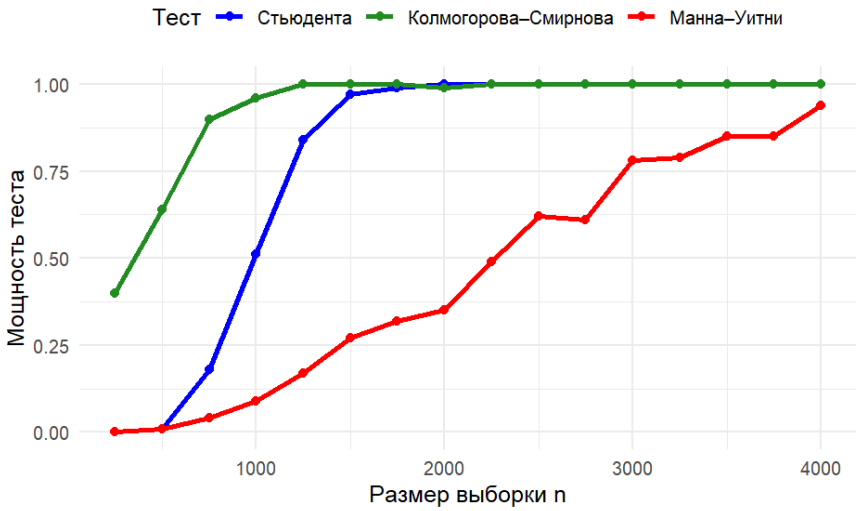
└Расстояние между распределением и его номинативным

Чтобы проверить устойчивость приближения распределения его номинативным аналогом, было проверено Как ведёт себя симметризованное информационное расстояние между распределением группы LOW и его номинативным, в зависимости от объема выборки.

Как видно из графика, для любых объемов выборки расстояние остается в пределах от 0.004 до 0.0044, и уже начиная с малых n колеблется вокруг среднего значения.

Чувствительность тестов к различиям в параметре β

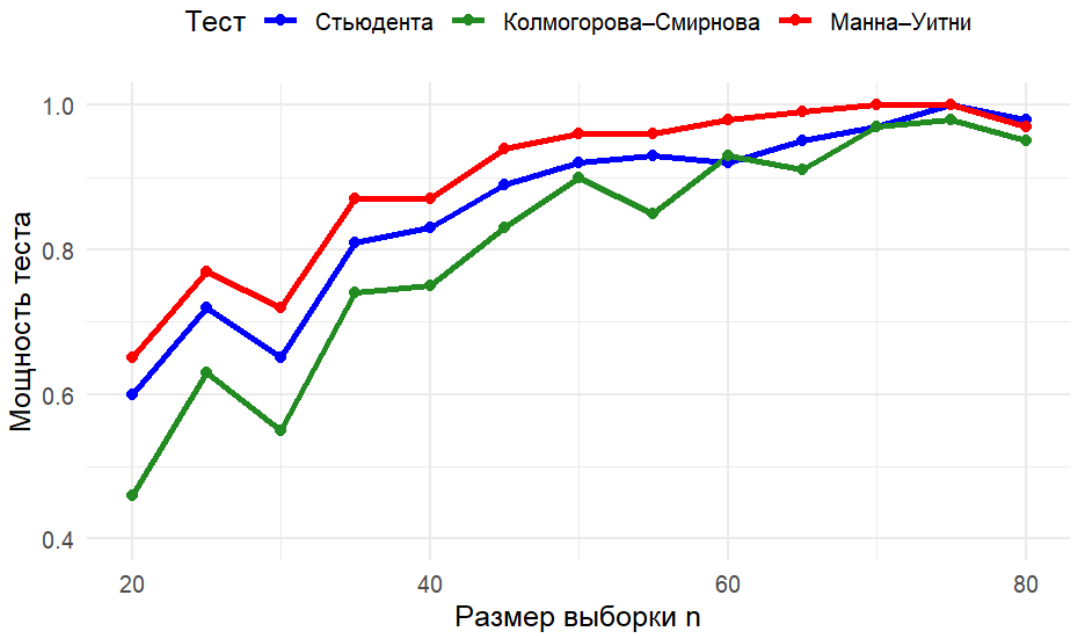
- Для 100 генераций с параметрами как в группах LOW и HIGH для каждого n.
- Мощность ≥ 0.84 достигается при:
  - $n \geq 1250$ для критерия Стьюдента;
  - $n \geq 750$ для Колмогорова–Смирнова;
  - $n \geq 3500$ для Манна–Уитни.

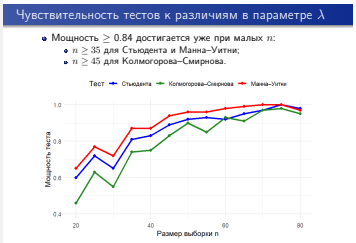


Следующим этапом работы было выяснить, какие объемы выборок требуются, чтобы найти значимые различия в параметрах распределения групп HIGH и LOW. Для этого была проведена проверка на модельных выборках.
Для каждого объема n сгенерировано 100 пар выборок с параметрами как в группах LOW и HIGH и оценивалась мощность трёх тестов.
На графике видно, что: • t-критерию требуется как минимум n = 1 250, • Колмогорову–Смирнову хватает n = 750, • а непараметрическому Манна–Уитни — почти n = 3 500.
При наших реальных объемах (446 и 3 554) не хватило для обнаружения значимого различия в параметре бета.

Чувствительность тестов к различиям в параметре λ

- Мощность ≥ 0.84 достигается уже при малых n:
  - $n \geq 35$ для Стьюдента и Манна–Уитни;
  - $n \geq 45$ для Колмогорова–Смирнова.





Таким же образом проверялась мощность критериев при проверке однородности параметра лямбда.
Здесь ситуация кардинально другая: мощность 0.84 достигается уже на очень маленьких объемах. • t-критерий и Манна–Уитни справляются при n=35, • Колмогоров–Смирнов — при n=45.
Это подтверждает результаты, полученные при проверке однородности параметра лямбда, различия значимы.

Регрессия с гамма-распределенной ошибкой

Модель:

$$y_i = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon_i, \quad \varepsilon_i \sim \Gamma(\lambda, \beta)$$

- y_i – ИМТ.
- x_i – TAS (алекситимия), EDE и DEBQ (опросники по РПП), возраст.
- 20 выборок объемом $n = 200$.
- Для каждой переменной x_i в проверялась однородность параметров гамма-распределения остатков между полной моделью с 4 предикторами и укороченной моделью (без этой переменной, с 3 предикторами).

Статистические модели на основе гамма распределения

└Регрессия с гамма-распределенной ошибкой

Регрессия с гамма-распределенной ошибкой

Модель:

$$y_i = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon_i, \quad \varepsilon_i \sim \Gamma(\lambda, \beta)$$

- y_i – ИМТ.
- x_i – TAS (алекситимия), EDE и DEBQ (опросники по РПП), возраст.
- 20 выборок объемом $n = 200$.
- Для каждой переменной x_i в проверялась однородность параметров гамма-распределения остатков между полной моделью с 4 предикторами и укороченной моделью (без этой переменной, с 3 предикторами).

Следующим этапом моей работы было построение модели с гамма распределенной ошибкой.
В качестве зависимой берется ИМТ, в качестве независимых — три психологических опросника: TAS (Алекситимия · затруднение в определении и описании (вербализации) собственных эмоций и эмоций других людей;), EDE и DEBQ, возраст.
Чтобы оценить вклад каждой переменной, я беру 20 подвыборок по 200 человек и для каждой сравниваю пару моделей: • полную — со всеми четырьмя предикторами; • укороченную — в которой отсутствует данная переменная.
Затем сравниваю распределения остатков двух моделей по параметрам гамма-распределения. Если при исключении переменной параметры лямбда или бета меняются значимо, значит эта переменная действительно влияет на форму распределения ИМТ.

Влияние переменных на параметры распределения остатков

Table: Количество выборок (из 20 по 200 человек), в которых исключение переменной привело к значимому различию в параметрах гамма-распределения остатков ($\alpha = 0.05$, критерий Стьюдента).

Переменная	λ **(форма)**	β **(масштаб)**
TAS (алекситимия)	1	0
EDE (РПП)	9	0
DEBQ (РПП)	1	0
Возраст	**18**	0

Влияние переменных на параметры распределения остатков

Table: Количество выборок (из 20 по 200 человек), в которых исключение переменной привело к значимому различию в параметрах гамма-распределения остатков ($\alpha = 0.05$, критерий Стьюдента).

Переменная	λ **(форма)**	β **(масштаб)**
TAS (алекситимия)	1	0
EDE (РПП)	9	0
DEBQ (РПП)	1	0
Возраст	**18**	0

Таблица показывает, в скольких из 20 подвыборок исключение каждой переменной вызывало значимое изменение параметров гамма-распределения остатков для каждой переменной. Сравниваем по критерию стьюдента, по методике, которая была описана ранее. Видим, что в параметре масштаба нет значимых отличий ни для одной переменной. А вот параметр формы реагирует:

• Возраст — самый сильный фактор: в 18 из 20 выборок исключение возраста делает распределение остатков уже или шире. • Тяжесть РПП по шкале EDE влияет только в примерно половине случаев.

Значит, вариативность ИМТ прежде всего формируется возрастом и тяжестью расстройства, тогда как масштаб остаётся устойчивым.

Заключение: интерпретация результатов

- Различие между группами LOW (норма) и HIGH (риск РПП):
  - параметр формы λ: большая неоднородность в группе HIGH => большее количество факторов, влияющих на ИМТ (психологических, поведенческих, генетических).
- Модель с гамма-распределенными остатками:
  - Возраст влияет на форму распределения остатков (параметр формы λ), что отражает возрастную динамику в физиологии и пищевом поведении.

Статистические модели на основе гамма распределения

└Заключение: интерпретация результатов

Заключение: интерпретация результатов

- Различие между группами LOW (норма) и HIGH (риск РПП):
  - параметр формы λ: большая неоднородность в группе HIGH => большее количество факторов, влияющих на ИМТ (психологических, поведенческих, генетических).
- Модель с гамма-распределенными остатками:
  - Возраст влияет на форму распределения остатков (параметр формы λ), что отражает возрастную динамику в физиологии и пищевом поведении.

Первое — сравнение групп LOW и HIGH.
Мы показали, что масштаб можно считать однородным, а вот параметр формы значимо больше. Это можно обьяснить тем, что: у женщин с высоким риском РПП распределение ИМТ более неоднородное, большее количество факторов: генетика, психология, поведение, среда.
Регрессия с гамма распределенными остатками. Сильный вклад в форму распределения остатков даёт возраст: по мере старения, накапливается влияние факторов метаболизма, привычек, генетической предрасположенности.

Заключение: результаты работы

- Предложен подход к сравнению групп с помощью проверки однородности параметров гамма-распределения.
- Использованы синонимичные распределения для расширения подхода.
- Проведён анализ мощности статистических критериев.
- Установлена устойчивость приближения: расстояние J между распределением и номинативной моделью стабильно мало при любом объёме выборки.
- Построена регрессионная модель с гамма-распределённой ошибкой; проанализирован вклад переменных в структуру остатков.

Статистические модели на основе гамма распределения

└Заключение: результаты работы

Заключение: результаты работы

- Предложен подход к сравнению групп с помощью проверки однородности параметров гамма-распределения.
- Использованы синонимичные распределения для расширения подхода.
- Проведён анализ мощности статистических критериев.
- Установлена устойчивость приближения: расстояние J между распределением и номинативной моделью стабильно мало при любом объёме выборки.
- Построена регрессионная модель с гамма-распределённой ошибкой; проанализирован вклад переменных в структуру остатков.

В заключение, кратко обозначу результаты моей работы.
• Была предложена методика сравнения групп с помощью проверки однородность параметров масштаба и формы гамма-распределения.
• Далее, использовались синонимичные распределения, чтобы обобщить поход, подтвердить полученные результаты
• Кроме этого, проведена оценка мощности тестов, результаты показали что различия в параметре формы обнаруживаются уже при малых эн, тогда как бета требует больших выборок.
• Также была показана устойчивость приближения распределения его номинативным аналогом: результаты показывают, что информационное расстояние J между исходным и номинативным распределением остаётся малым при любом объёме выборки.
• И наконец, построили гамма-регрессию и выявили, что возраст сильно влияет на форму распределения ИМТ.
На этом мой доклад подходит к концу. Благодарю за внимание!