



Факультет гуманитарных наук

Школа
лингвистики

Москва
2025

Исследование графических маркеров юмора в неформальной интернет-переписке

Меньшова Полина Максимовна



Репозиторий проекта:



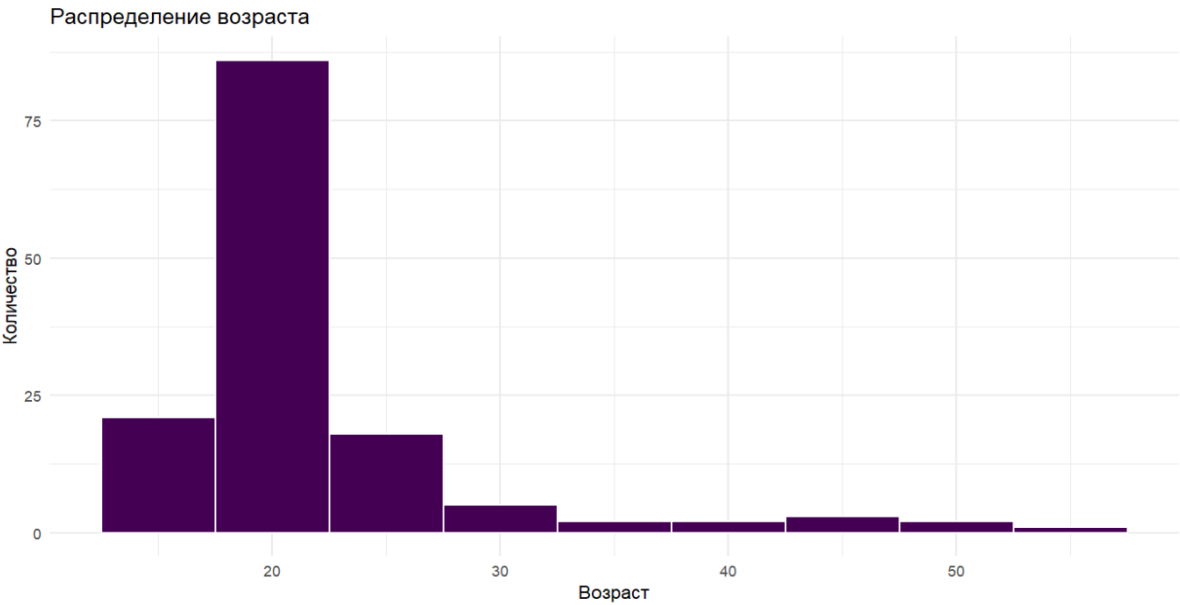
Обзор релевантных исследований

- Лексические средства > Графические средства (Гуторина, 2019)
- Автоматическое распознавание иронии (Kosterin et al, 2023)
- Спектр контекстуальной информации (Yus, 2023)
- Ирония vs сарказм (Katz, 2023)
- Оценка способности человека распознавать иронию (Bruntsch & Ruch, 2017)



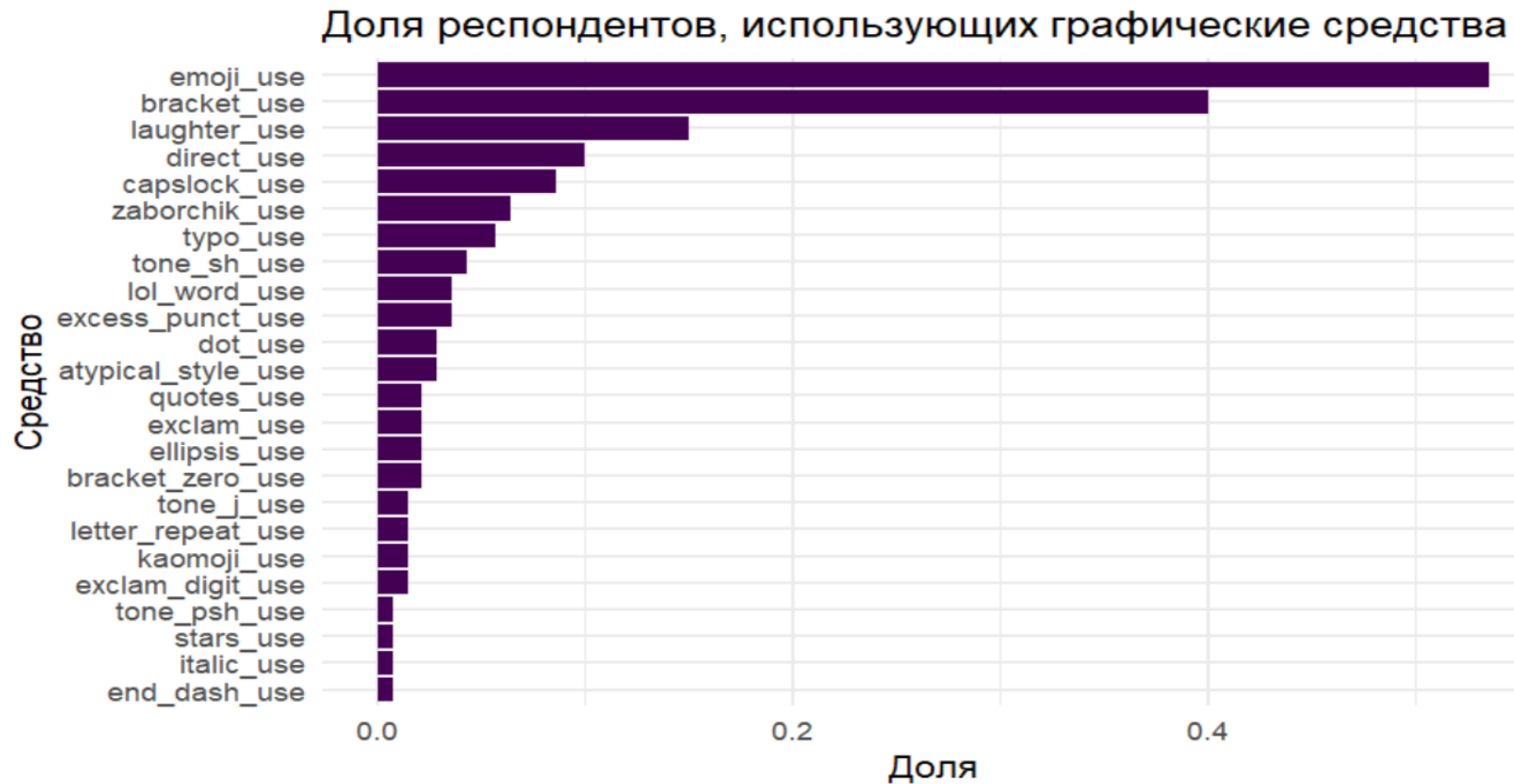
Разведочный анализ

min_age	max_age	mean_age	median_age	sd_age
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
13	55	22.23571	21	6.9888



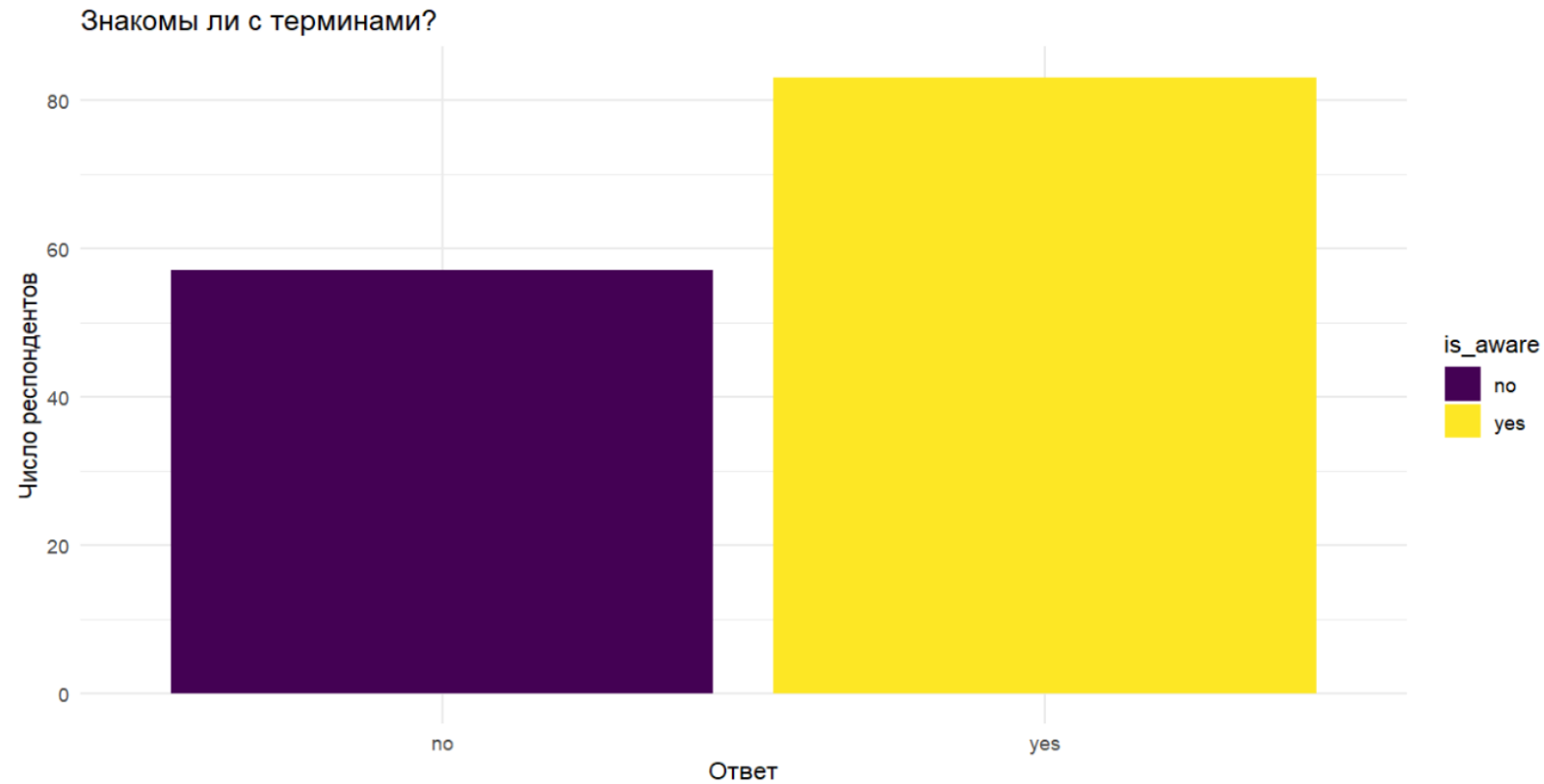


Разведочный анализ



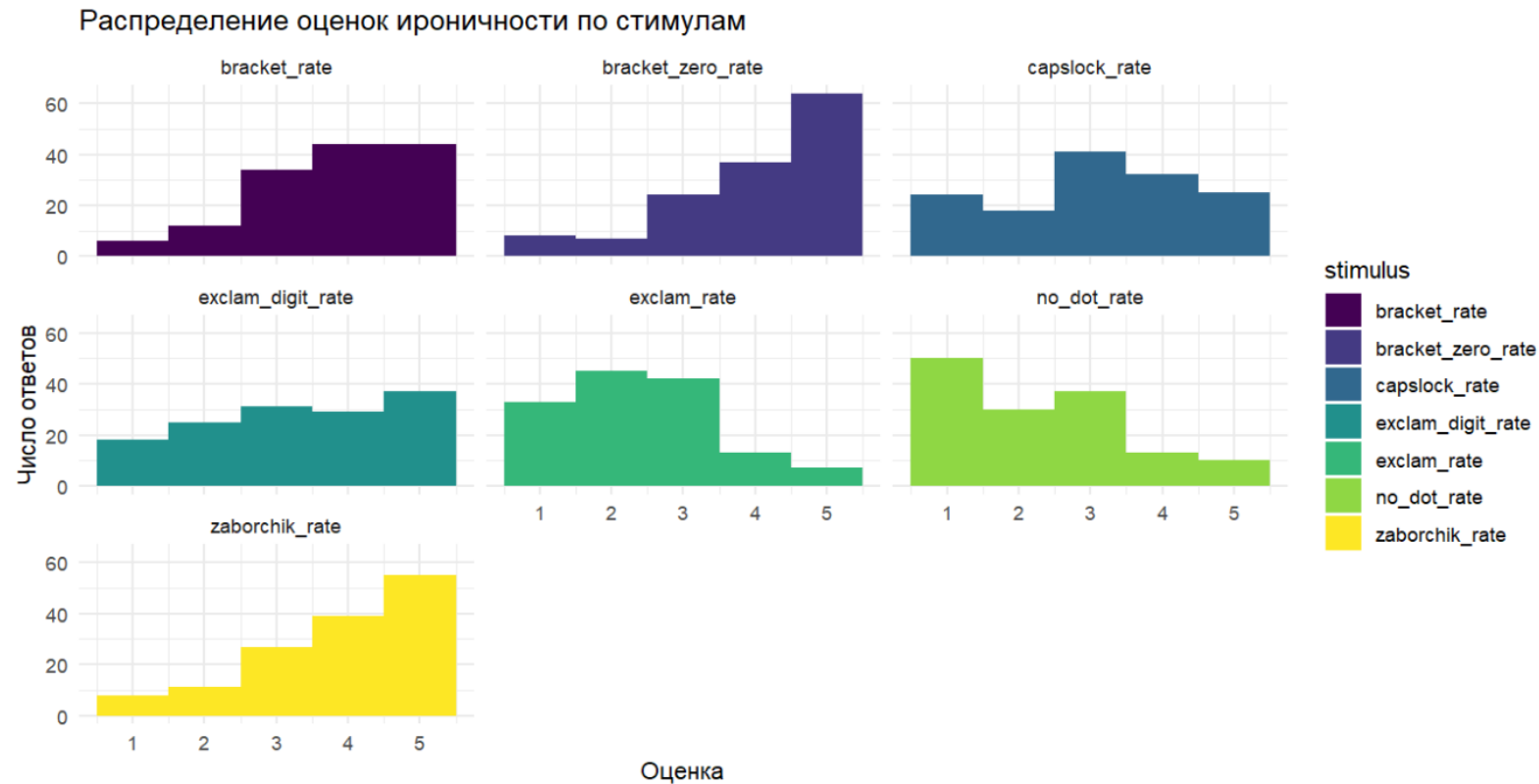


Разведочный анализ





Разведочный анализ





Нулевая гипотеза

H_0 : Респонденты, использующие оцениваемые графические средства в своих сообщениях (bracket_use = "yes" и т. д.), **не ставят** более высокие оценки ироничности стимулам с этими средствами (bracket_rate и т. д.).

Методы

- Тест Уилкоксона
- Коэффициент корреляции Спирмена
- Сравнение средних
- Анализ мощности тестов
- Множественная регрессия (in question)



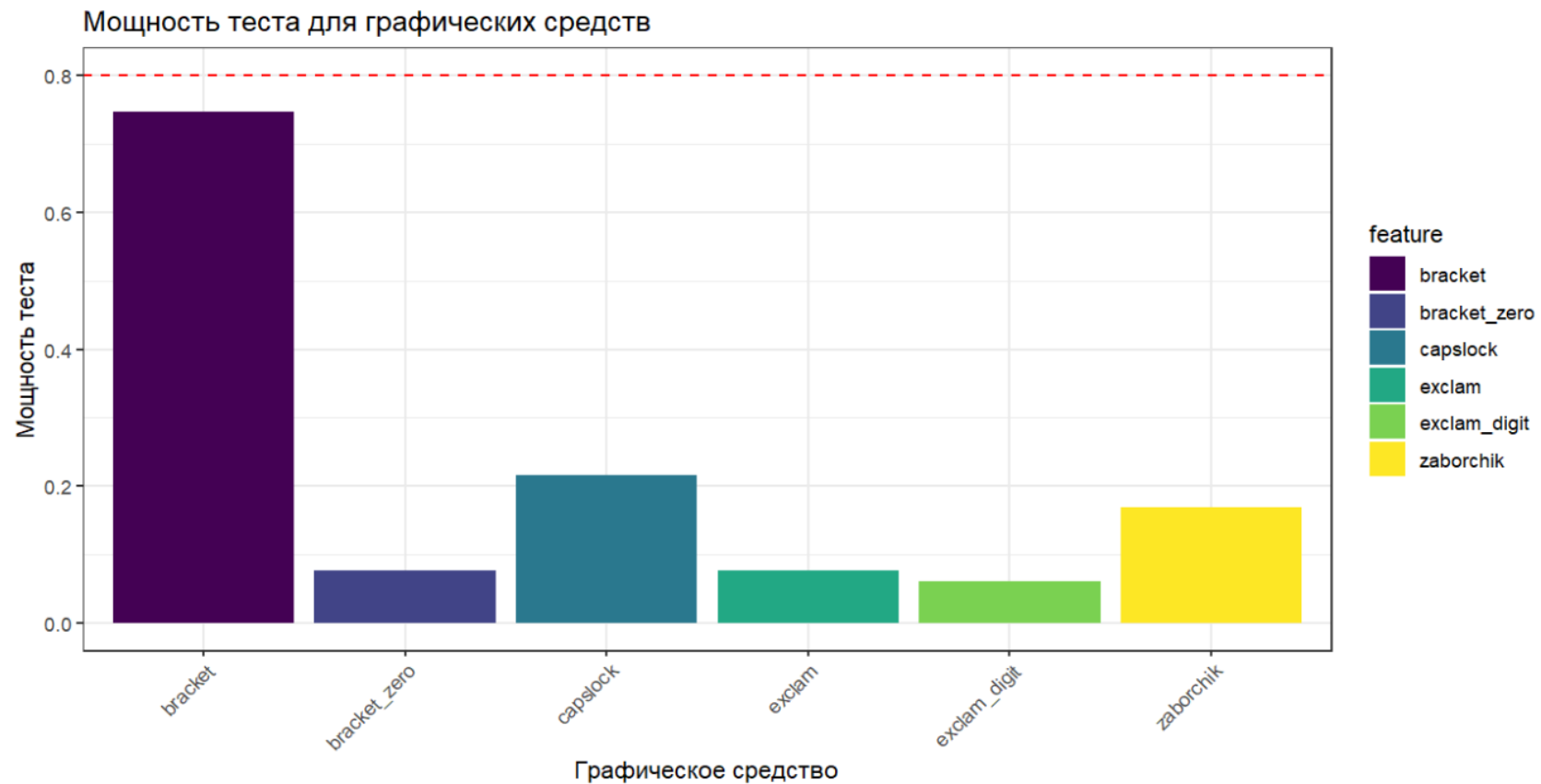
Тест Уилкоксона, корреляция Спирмена и сравнение средних

- Отвергаем H_0 для скобок и капслова
- Не отвергаем H_0 для скобок с нулями, заборчика, восклицательных знаков и восклицательных знаков с единицами

Стимул	Тест Уилкоксона	Коэффициент корреляции Спирмена	Среднее (да)	Среднее (нет)
Скобки	$p = 0.048$	$\rho = 0.168$ слабый эффект ($p = 0.047$)	4.02	3.61
Капслок	$p = 0.033$	$\rho = 0.181$ умеренная связь ($p = 0.032$)	3.92	3.04



Анализ мощности тестов





Множественная регрессия с ковариатами

- Анализ влияния факторов помимо использования графических средств
- Уменьшение дисперсии остатков
- Увеличение статистической мощности (и точности оценок)
- Может быть неэффективна для дискретной величины



Множественная регрессия с ковариатами

$\text{rating} \sim \text{use_numeric} * \text{stimulus} + \text{age_scaled} + \text{is_aware_numeric}$

- *rating* — оценка ироничности
- *use_numeric* — индикатор использования графического средства (0/1)
- *stimulus* — тип графического стимула
- *age_scaled* — стандартизованный возраст
- *is_aware_numeric* — знание терминов (0/1)



Множественная регрессия с ковариатами

- Модель объясняет 18-20% вариации оценок ироничности
- Модель в целом статистически значима
- Все взаимодействия незначимы ($p\text{-value} > 0.05$), хотя для `use_numeric:stimulusexclam` наблюдается тенденция ($p\text{-value} = 0.08$)

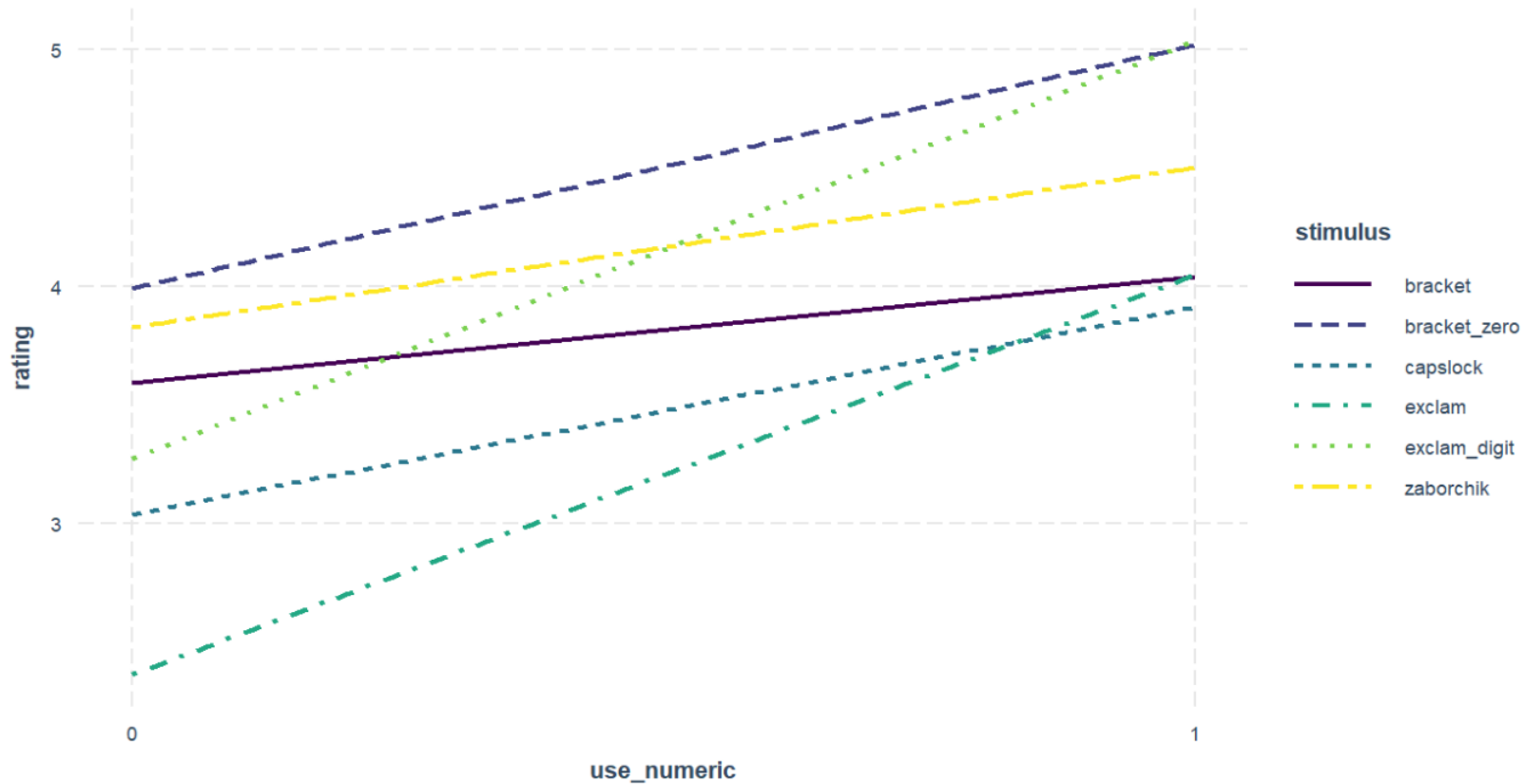
Residuals:

Min	1Q	Median	3Q	Max
-3.0850	-0.9208	0.0584	1.0017	2.6993

Residual standard error: 1.2 on 826 degrees of freedom
Multiple R-squared: 0.1997, Adjusted R-squared: 0.1871
F-statistic: 15.85 on 13 and 826 DF, $p\text{-value} < 2.2e-16$



Множественная регрессия с ковариатами





Выводы

- Использование определённых графических средств **действительно связано** с оцениванием того или иного средства как более или менее явно указывающего на иронию
- Тесты относительно мощные **только для скобок**
- Множественная регрессия показывает **тенденцию для капслока**, но её не стоит считать высокоэффективным методом анализа для имеющихся данных
- **Скобки и капслок** при имеющихся данных на выбранном уровне значимости можно назвать графическими маркерами иронии в неформальной переписке (по крайней мере для людей 18-25 лет)



Дискуссия и ограничения

- **Выборка:** больше данных, репрезентативность по возрасту и использованию графических средств
- **Дизайн проекта:** рандомизировать порядок демонстрации стимулов респондентам, добавить контрольную группу или группы, ограничить время экспозиции
- **Методы:** eye-tracking, интервью, стандартизированный тест для оценки способности распознавать иронию
- **Терминология:** ирония и сарказм, ироничная похвала и ироничная критика



Полина Меньшова
polina.menshova.19@mail.ru