

**Семинар НУГ**

**“Проверка устойчивости  
социального почерка  
на логах коммуникации в Telegram”**

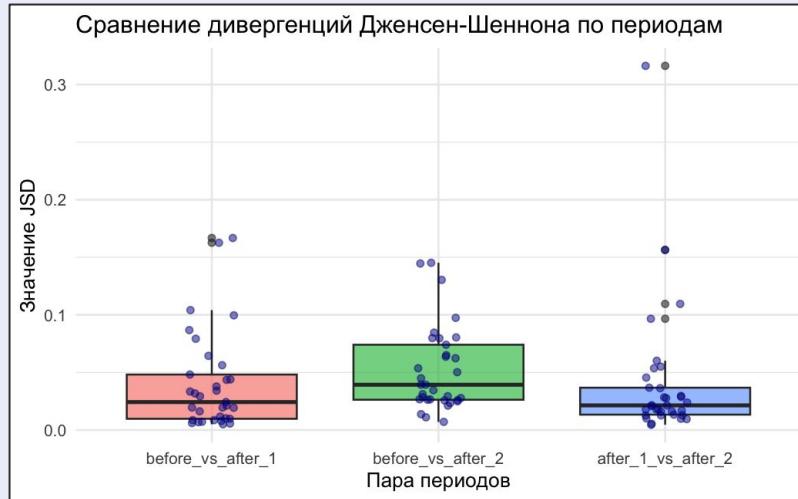
## Проверка корректности подсчётов

Выбираем трёх рандомных девушек (set.seed) и для каждой из них проверяем, правильно ли последовательно преобразуется датасет

- как посчитались бины по выбранным периодам (верно ли определены границы периода в зависимости от заданной даты рождения ребёнка, правильно ли преобразована временная метка в бин)
- как проранжированы альтеры по количеству бинов...

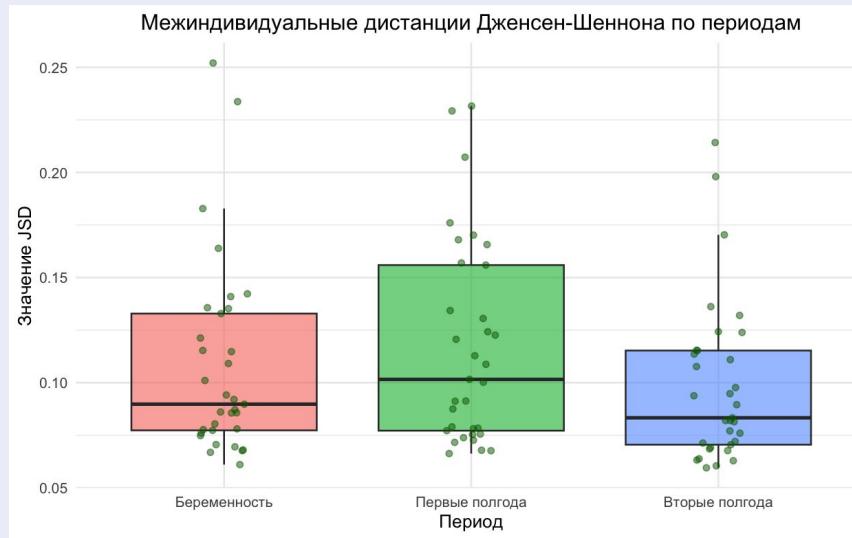
**N = 33** женщины

**Расчёт внутрииндивидуальных дистанций Дженсен-Шеннона (JSD):**  
три периода по полгода (до рождения ребёнка, первые и вторые полгода после),  
**бин = 1 час, ограничение в 20 альтеров**



Среднее значение дистанции между 1 и 2 периодом ( $JSD_{self\_1\_2}$ ) равно 0,040  
Между 1 и 3 периодом ( $JSD_{self\_1\_3}$ ) равно 0,052  
Между 2 и 3 периодом ( $JSD_{self\_2\_3}$ ) равно 0,041

**Расчёт межиндивидуальных дистанций Дженсен-Шеннона (JSD):  
среднее значение дистанции между распределением одной женщины  
и распределениями всех остальных женщин в выборке в рамках одного периода**



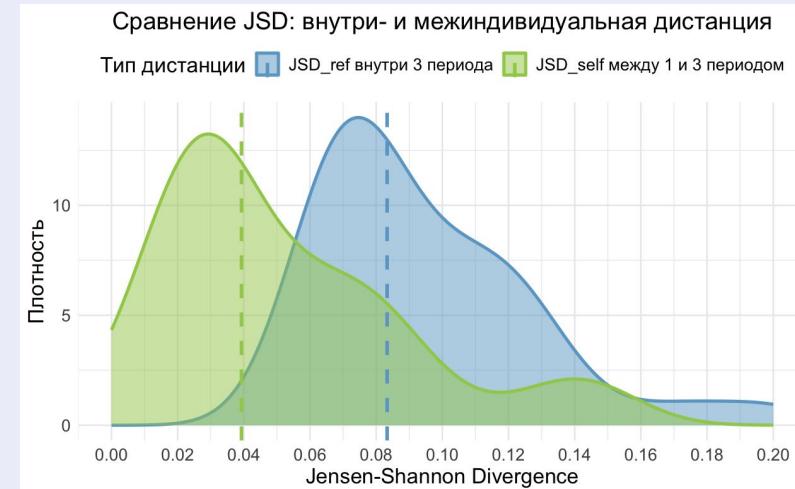
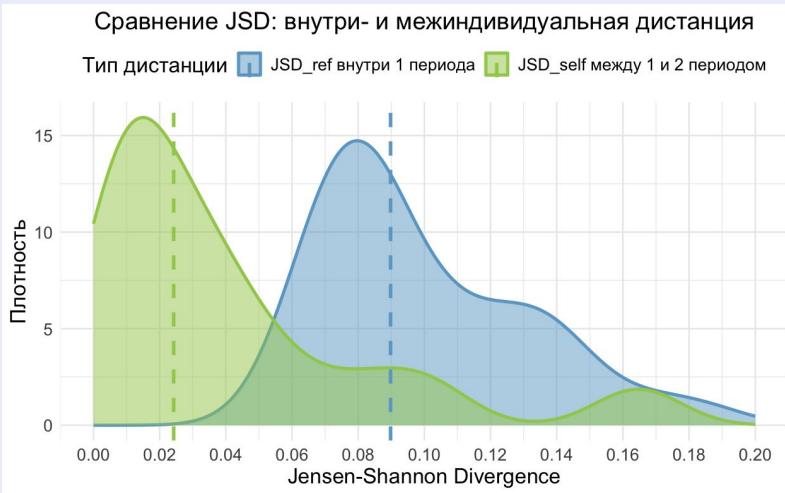
**Среднее значение дистанции в рамках первого периода (JSDef\_before) равно 0,108  
В рамках второго периода (JSDef\_after\_1) равно 0,117  
В рамках третьего периода (JSDef\_after\_2) равно 0,098**

**Гипотеза: Дистанция между социальным почерком одной и той же женщины (внутрииндивидуальная) до и после рождения ребенка меньше, чем дистанция между социальным почерком разных женщин (межиндивидуальная) в один и тот же период их материнства.**

Дистанция JSDself	Дистанция JSDref	Среднее JSDself	Среднее JSDref	W-значение ( $p<0,01^{**}$ )
JSDself между 1 и 2 срезом	JSDref внутри 1 среза	0,040	0,108	16**
JSDself между 1 и 2 срезом	JSDref внутри 2 среза	0,040	0,117	28**
JSDself между 2 и 3 срезом	JSDref внутри 2 среза	0,041	0,117	41**
JSDself между 2 и 3 срезом	JSDref внутри 3 среза	0,041	0,098	53**
JSDself между 1 и 3 срезом	JSDref внутри 1 среза	0,052	0,108	26**
JSDself между 1 и 3 срезом	JSDref внутри 3 среза	0,052	0,098	59**

Во всех 6 тестах Уилкоксона (Манна-Уитни для зависимых выборок) **внутрииндивидуальные дистанции оказались значимо меньше межиндивидуальных** ( $JSDself < JSDref$ ,  $p<0,01$ ), что указывает на схожесть распределений одной и той же респондентки во времени, устойчивость её паттерна коммуникации.

# Визуализация различий



# **Анализ влияния выбора параметров социального почерка на вывод о его устойчивости**

Для оценки чувствительности социального почерка к параметрам были реализованы 16 комбинаций, которые различаются по трём параметрам:

**размер временного бина:**

- 1 час,
- 1 день;

**число включаемых альтеров:**

- 15, 20, 40, 45 наиболее активных контактов;

**длина временного периода:**

- полгода,
- три месяца.

Схемы построения временных периодов:

## **1. Полугодовые интервалы, привязанные к дате рождения ребенка:**

Первый интервал - 6 месяцев до рождения ребенка

Второй интервал - 6 месяцев после рождения ребенка

Третий интервал - дополнительные 6 месяцев после рождения ребенка (в сумме 12 месяцев после рождения ребенка)

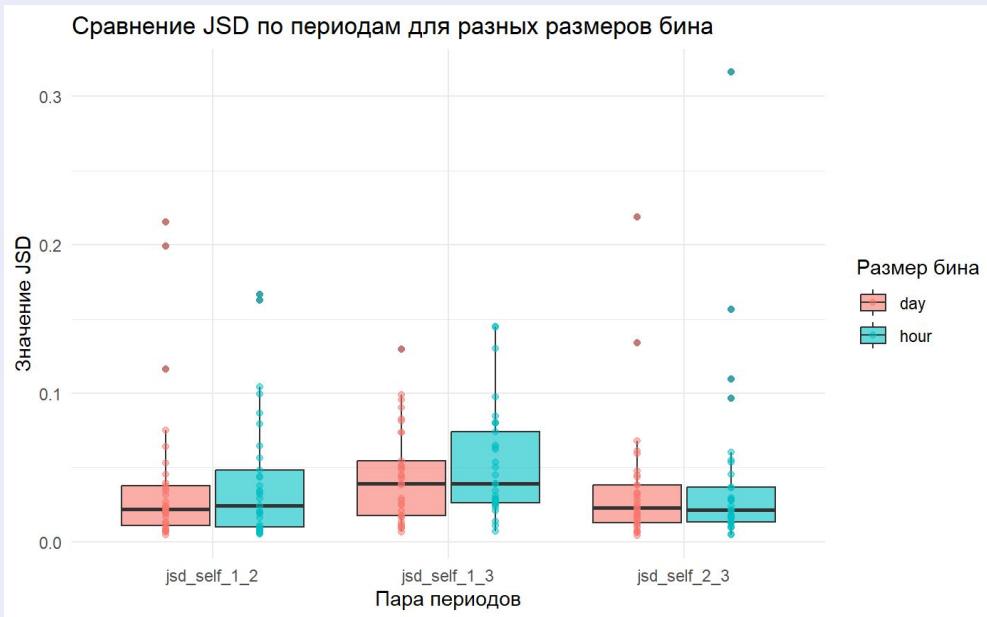
## **2. Трехмесячные интервалы, привязанные к дате рождения ребенка:**

Первый интервал - 3 месяца до рождения ребенка

Второй интервал - 3 месяца после рождения ребенка

Третий интервал - дополнительные 3 месяца после рождения ребенка (в сумме 6 месяцев после рождения ребенка)

# Проверка влияния изменения бина



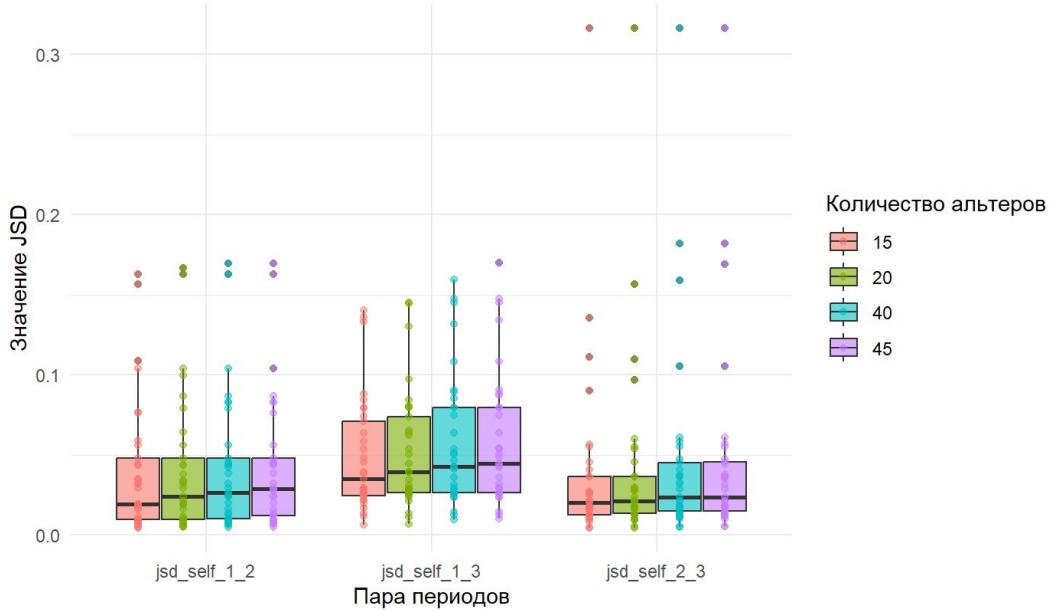
Исходная модель для проверки влияния изменчивости параметров: 20 альтеров, бин равный 1 часу и полугодовые интервалы, привязанные к дате рождения ребенка.

## Результаты Wilcoxon-теста:

- jsd\_self\_1\_2,  $p = 0.49$
- jsd\_self\_1\_3,  $p = 0.23$
- jsd\_self\_2\_3,  $p = 0.44$

# Проверка влияния изменения количества альтеров

Сравнение JSD по периодам для разных значений количества альтеров



top_k1 <dbl>	top_k2 <dbl>	period_pair <fctr>	p_value <dbl>
15	20	jsd_self_1_2	0.0464686337
15	40	jsd_self_1_2	0.0005729962
20	40	jsd_self_1_2	0.0010416431
15	45	jsd_self_1_2	0.0005729962
20	45	jsd_self_1_2	0.0011680676
40	45	jsd_self_1_2	0.3065562734
15	20	jsd_self_1_3	0.0032498293
15	40	jsd_self_1_3	0.0001543667
20	40	jsd_self_1_3	0.0001532677
15	45	jsd_self_1_3	0.0000465637
20	45	jsd_self_1_3	0.0001216103
40	45	jsd_self_1_3	0.0025844789
15	20	jsd_self_2_3	0.0012573390
15	40	jsd_self_2_3	0.0003769499
20	40	jsd_self_2_3	0.0053864082
15	45	jsd_self_2_3	0.0002088163
20	45	jsd_self_2_3	0.0027366048
40	45	jsd_self_2_3	0.0445533169

# Проверка влияния изменения размера временного интервала

Результаты Wilcoxon-теста:

- jsd\_self\_1\_2, p = 0.32
- jsd\_self\_1\_3, p = 0.26
- jsd\_self\_2\_3, p = 0.96

