

# Варьирование несклоняемых существительных по грамматическому роду

# Гипотеза

- H0: носители русского языка определяют грамматический род несклоняемых существительных на основе **конечного звука**.
- H1: носители русского языка определяют грамматический род несклоняемых существительных на основе **другой характеристики**.

# Материалы исследования

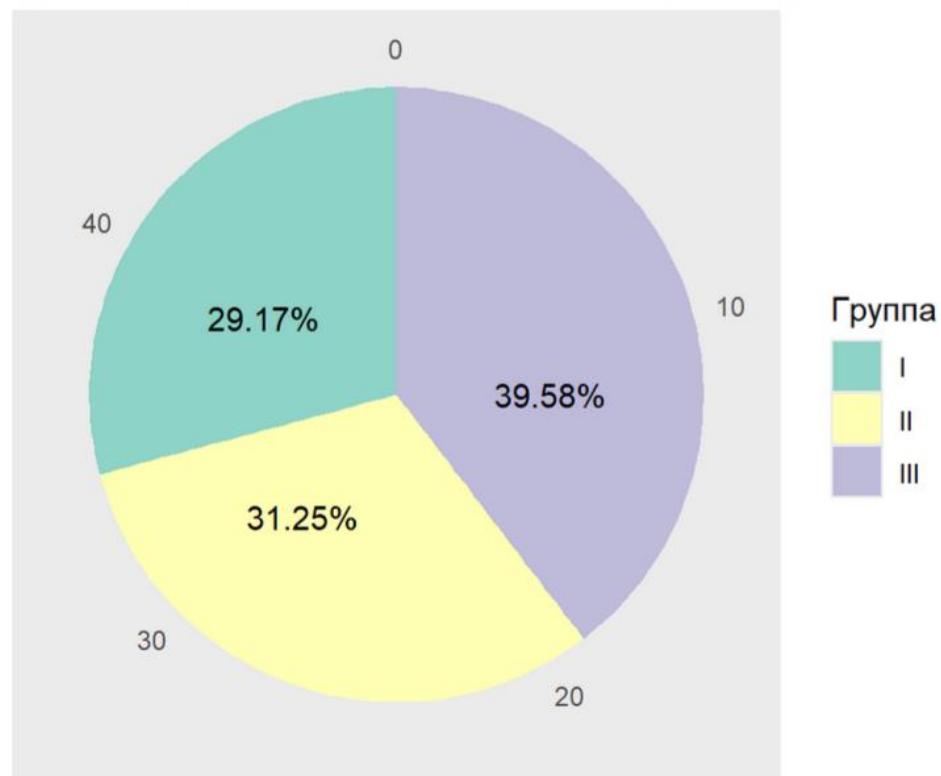
## Перечень существительных:

- 48 несклоняемых существительных
- Группа существительных: I (неологизм), II (частое), III (редкое)
- Конечный звук

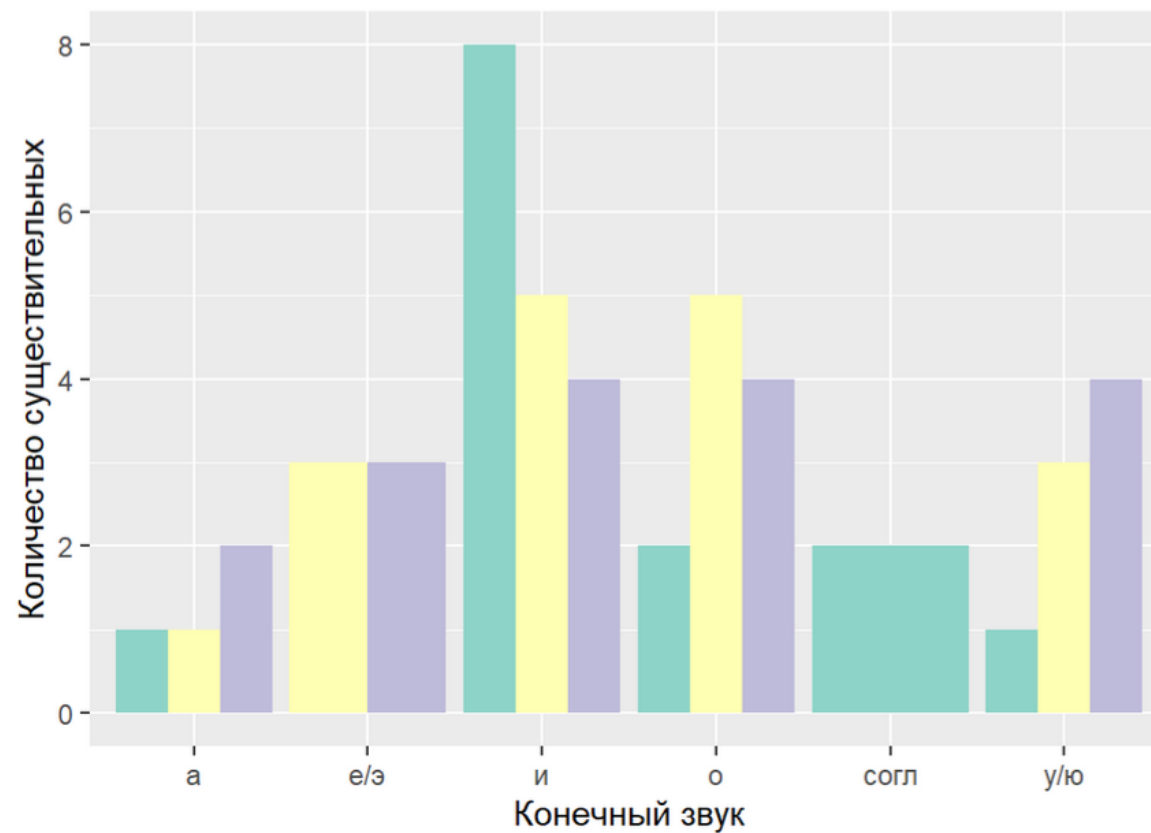
	A	B	C	D	E
1	word	group	gram_gender	origin	last_letter
10	моти	I	n*	японский*	и
11	демо	I	-	английский*	о
12	гуаша	I	-	китайский*	а
13	барбекю	II	n	французский	у/ю
14	табу	II	n	французский	у/ю
15	шоу	II	n	английский	у/ю
16	акне	II	f,n	французский	е/э
17	кофе	II	m,n	голандский	е/э

# Материалы исследования

Группы существительных после пересмотра



Группы существительных и конечные звуки



# Материалы исследования

## Опросник:

- 48 несклоняемых существительных (=предложений)
- Частота использования слова
- Уверенность в понимании смысла слова

Ниже представлено предложение со словом ХОББИ. Вам необходимо выбрать окончание(-я) для пропусков и оценить, как часто вы употребляете это слово в своей речи.

Как бы вы сформулировали следующую фразу? \*

Таксидермия — [мо\_\_ нов\_\_ хобби].

☒ моё новое хобби

☐ мой новый хобби

☐ моя новая хобби

☐ Other: \_\_\_\_\_

Насколько хорошо вы знакомы со значением слова ХОББИ? \*

1 — не знаком(-а) вовсе

5 — прекрасно представляю значение и употребляю это слово в речи



Как часто вы употребляете слово ХОББИ в устной или письменной речи? \*

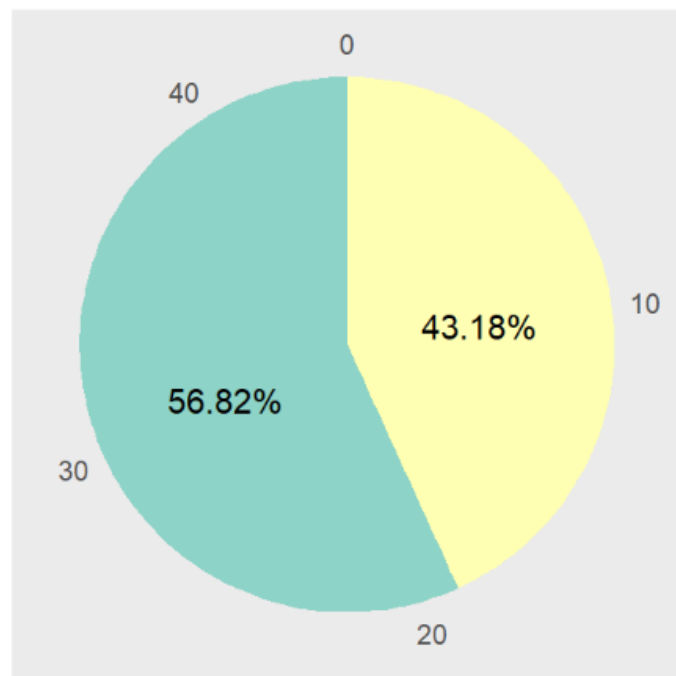
1 — не употребляю совсем

5 — употребляю это слово ежедневно



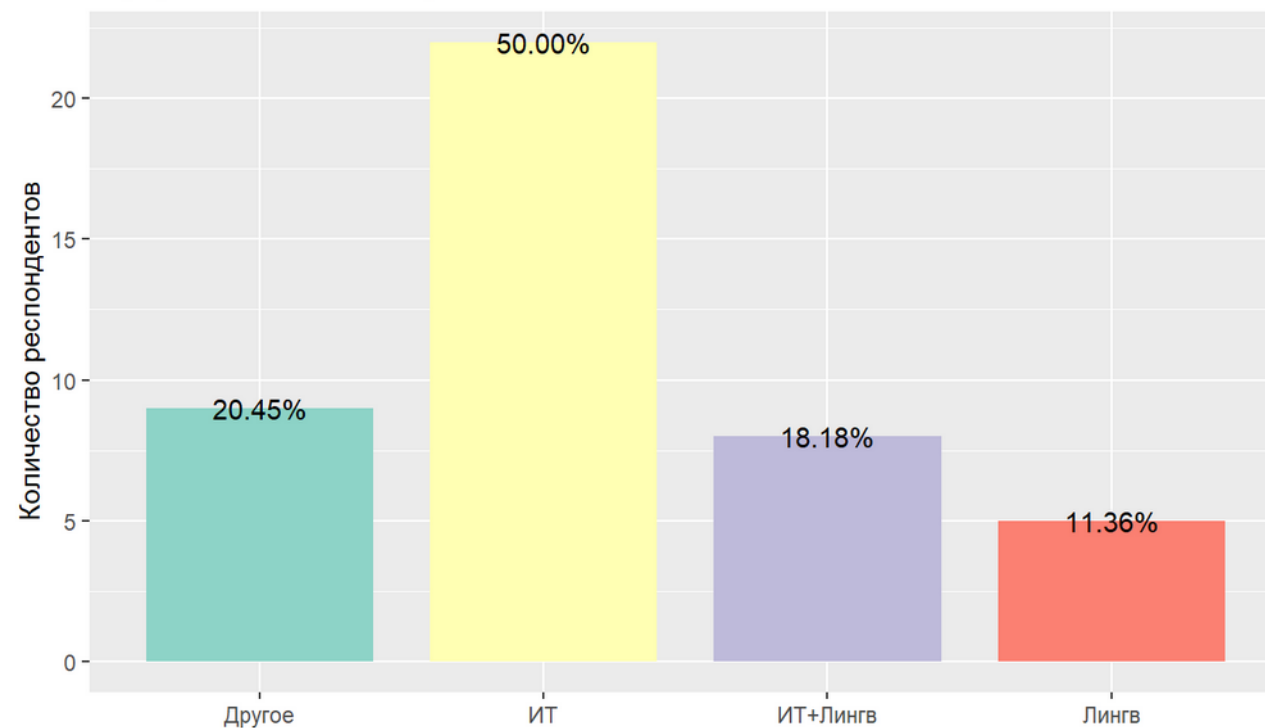
# Респонденты

Пол респондентов



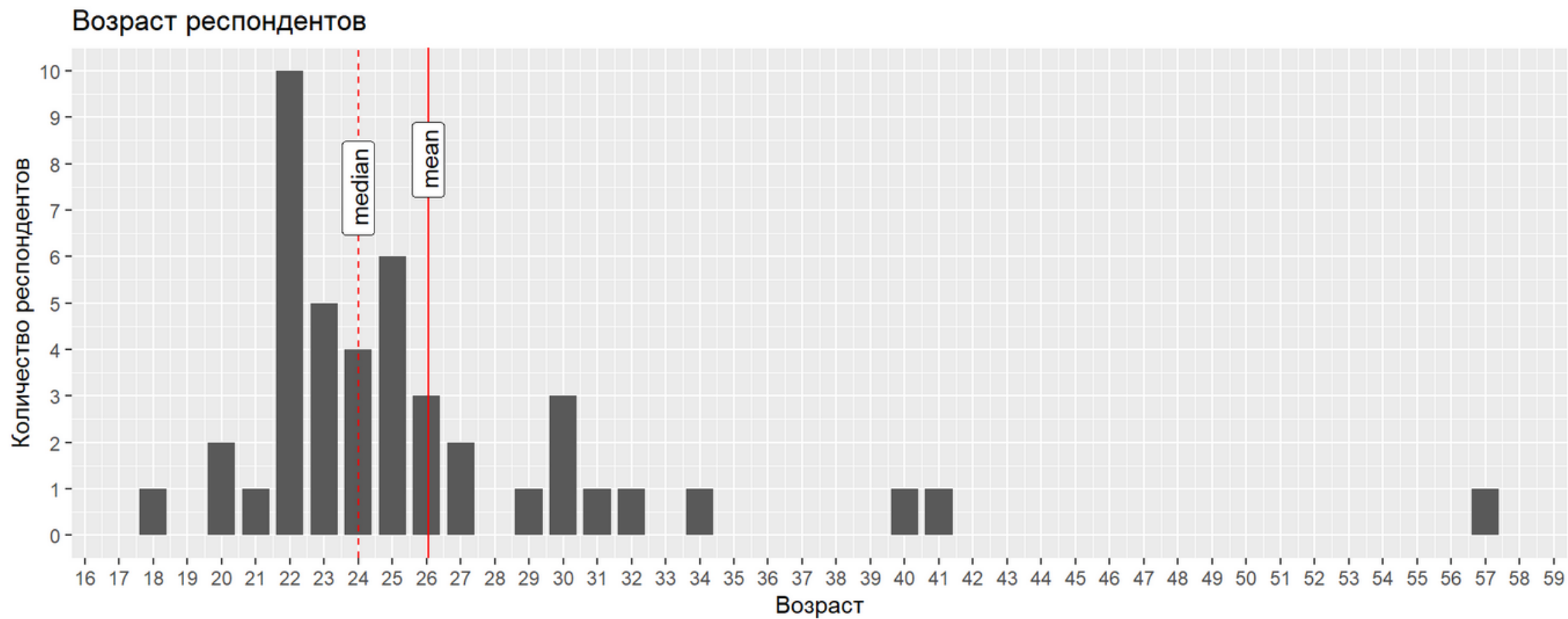
Пол  Женский  Мужской

Сфера деятельности респондентов



Сфера  Другое  ИТ  ИТ+Лингв  Лингв

# Респонденты





# Методы анализа

- Тест **Пирсона**: переменные грам. род - конечный звук
- однофакторный **ANOVA** с повторными измерениями: средняя энтропия ответов для слов с разными **конечными звуками**
- однофакторный **ANOVA** с повторными измерениями: средняя энтропия между **группами существительных**

# Тест Пирсона

```
# применяем тест Хи-квадрат Пирсона
chisq_ <- chisq.test(df_contig)
chisq_
```

```
##
## Pearson's Chi-squared test
##
## data:  df_contig
## X-squared = 285.46, df = 10, p-value < 2.2e-16
```

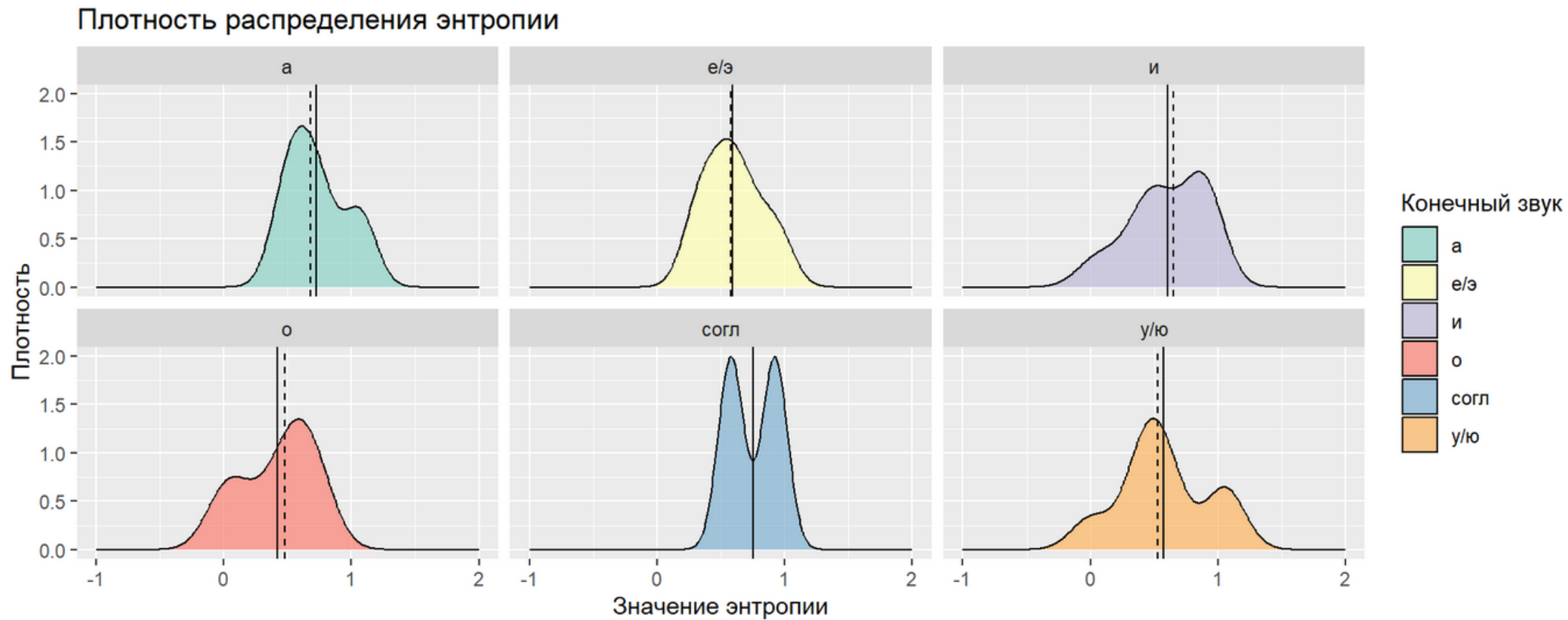
```
# считаем величину эффекта
effectsize::cramers_v(df_contig)
```

```
## Cramer's V (adj.) |          95% CI
## -----
## 0.25              | [0.22, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

Связь между переменными  
существует

Однако она не велика

# ANOVA: подготовка данных



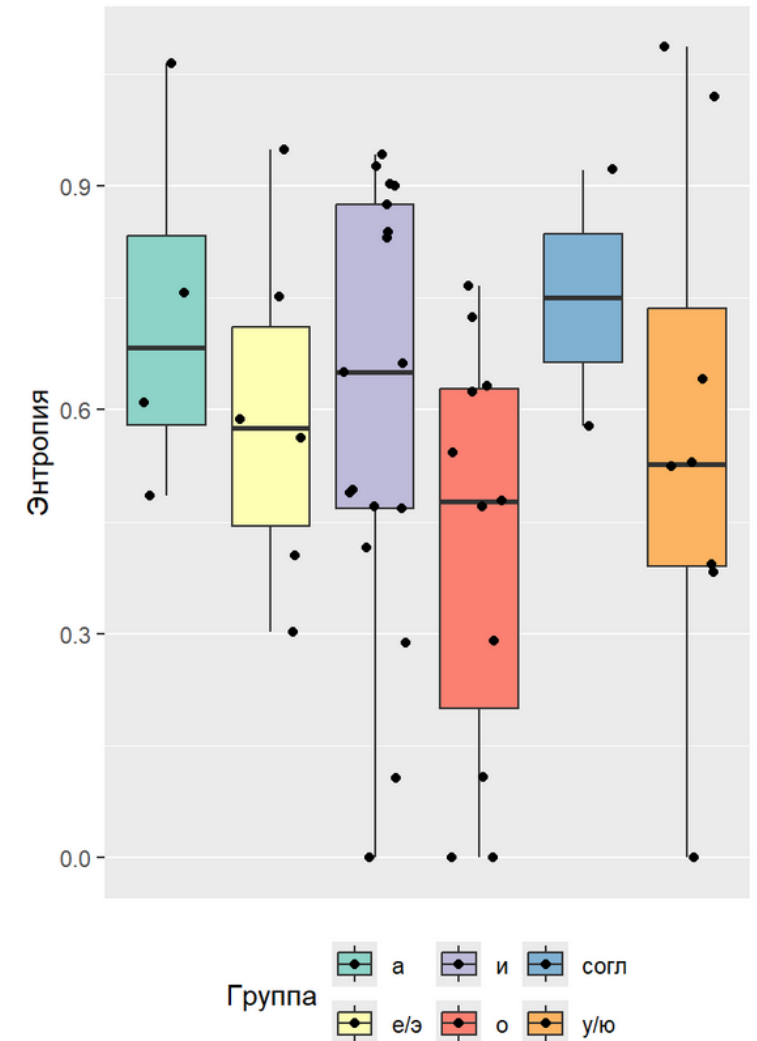
# Тест Фридмана: конечный звук – род

```
friedman_df <- df_sent2_grouped %>%  
  ungroup()  
  
model <- friedman.test(  
  friedman_df$entr, friedman_df$ParticipantID, friedman_df$last_lette  
r  
)  
  
model
```

```
##  
## Friedman rank sum test  
##  
## data:  friedman_df$entr, friedman_df$ParticipantID and friedman_df  
$last_letter  
## Friedman chi-squared = 59.124, df = 43, p-value = 0.05163
```

p-value > 0.05 => значительной разницы не существует

Энтропия и конечный звук

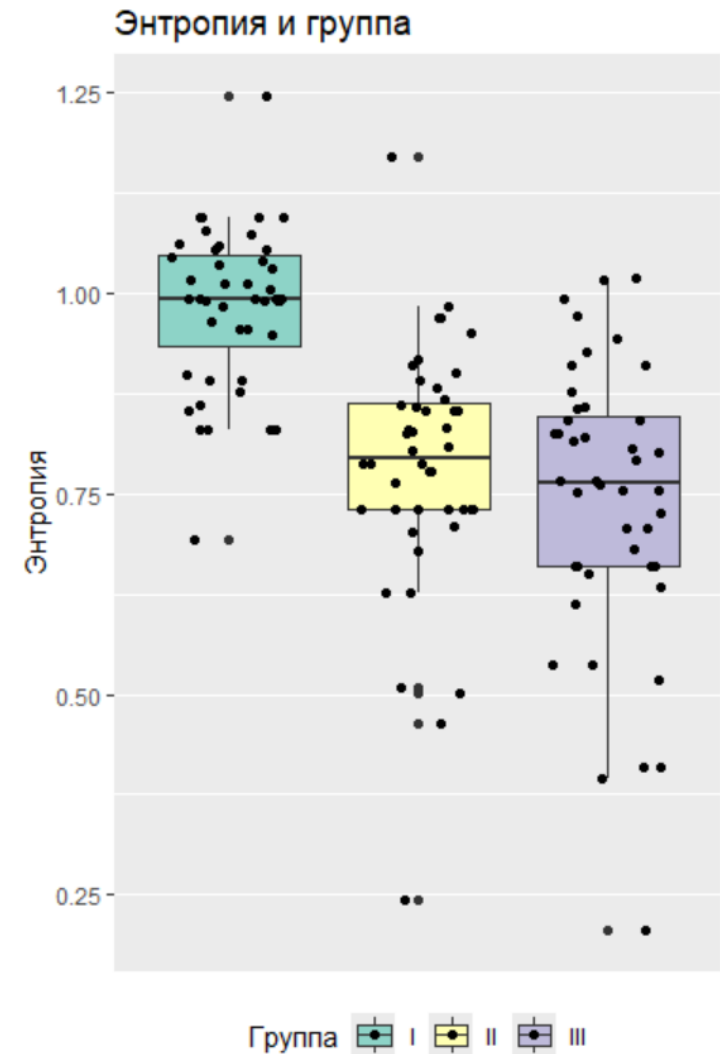


# Тест Фридмана: группа – род

```
friedman_df <- df_sent2_grouped %>%  
  ungroup()  
  
model <- friedman.test(  
  friedman_df$entr, friedman_df$ParticipantID, friedman_df$group_corr  
ected  
  )  
  
model
```

```
##  
## Friedman rank sum test  
##  
## data:  friedman_df$entr, friedman_df$ParticipantID and friedman_df  
$group_corrected  
## Friedman chi-squared = 55.476, df = 43, p-value = 0.09613
```

p-value > 0.05 => значительной разницы не существует



# Выводы

- Тест Пирсона подтвердил нулевую гипотезу о том, что определение грамматического рода несклоняемых существительных происходит с учетом конечного звука, однако стоит оговориться, что величина эффекта сравнительно мала.
- При этом ANOVA показал, что между словами, оканчивающимися на разные звуки, нет значимой разницы в энтропии, и ее значения находятся в пределах 0.5-0.8 – умеренная неопределенность.

# Ограничения

- Нерепрезентативная выборка
- Не совсем удачный выбор слов для группы I – неологизмы
- Ограниченный инструментарий методов анализа

# Вопросы и комментарии

1)

2)

3)