

Дата сдачи: 16.12. до 21.00

1. Зависит ли район проживания москвича от его национальности? [5 баллов]

Во многих мегаполисах есть кварталы, которые называются «китайский», «итальянский», «арабский». Иногда данные наименования носят исторический характер, иногда определяются преобладающим этническим составом населения квартала. Можно ли сказать про Москву, что район проживания человека не зависит от его национальности? К сожалению, данных этнического состава разных районов Москвы я не нашла, но представим такой упрощённый сценарий. (Для заинтересовавшихся: ссылка на этнический состав Москвы по переписи 2010г. <http://www.statdata.ru/nacionalnyj-sostav/moskvy>).

Сценарий

Предположим, что в Москве 4 округа: М, О, S, и С. Мы опросили 650 случайно выбранных жителей Москвы на предмет их национальной принадлежности (Russian, Tatar, Armenian). Нулевая гипотеза: округ проживания человека не зависит от его национальности. Данные представлены в таблице:

	М	О	S	С	Total
Russian	90	60	104	95	349
Tatar	30	50	51	20	151
Armenian	30	40	45	35	150
Total	150	150	200	150	650

Задание:

- 1 балл: приведите код, который Вы использовали для того, чтобы ввести данные в R.
- 1 балл: приведите код, который Вы использовали для подтверждения или опровержения нулевой гипотезы.
- 1 балл: подтверждается или опровергается нулевая гипотеза? Обоснуйте свой ответ.
- 2 балла: какой статистический тест Вы использовали для подтверждения или опровержения гипотезы? Обоснуйте свой выбор.

2. Статистическая шпаргалка [9 баллов]

Ваша собственная шпаргалка по статистическим критериям. Единственное требование – это объём в 9 статистических тестов. Формат – свободный, но шпаргалка должна быть достойна того, чтобы передать её по наследству.

Для каждого теста, Вы можете включить любую информацию, которая кажется Вам полезной и/или интересной: название, описание, пример данных, команды для R, ограничения на использование теста, упоминание статей, в которых он используется... Шпаргалка будет пополняться, поэтому я не ожидаю исчерпывающей информации по статистическим тестам. Шпаргалка может быть текстовым файлом, презентацией (можно даже вставить в неё слайды из моих презентаций с Вашими пояснениями) или, например, сохранённым рабочим файлом R с комментариями. Напоминание: комментарии в R начинаются с #.

Выполнение этого задания поможет подготовиться Вам к тесту 14.12.

3. Пилотный эксперимент по изучению противоречий [5 баллов]

Описание эксперимента

Experimental materials consisted of 15 sentences: 12 fillers and 3 target sentences. The experiment manipulated a single factor in the target sentences: whether the sentence is in a form of a regular embedding (49a), a sentence-internal shift (49b) or a sentence-final shift (49c). Target sentences were divided into three balanced lists, so that participants saw one and only one condition for each item. The example below presents target sentences from one of the three experimental lists:

- (49) a. Mary said that Sam is running for political office, but he isn't.
b. Katherine is, Dean thinks, getting married, but she isn't.
c. The hospital will hire a surgeon, Jeremy suspects, but it won't.

Filler sentences remained constant across experimental lists. Their function was to distract the participants from focusing solely on the target sentences and to create a landscape of (non)contradictory statements in order for the participants to fully appreciate the judgment scale. 12 fillers were of two types: 6 non-contradictory fillers (50) and 6 contradictory fillers (51).

- (50) a. Bob wants to go to the park, but Bill does not.
b. It's a good idea, but I can't endorse it.
c. Some, but not all, of the students passed the test.
d. John wants to go to a football game, but he won't.
e. James was planning to come to the party, but I don't think he will.
f. Sam thinks it will rain, and Joan thinks it will be sunny.
- (51) a. Ed has no children, and his youngest daughter is visiting him.
b. Judy stopped smoking, but she did not smoke before.
c. Jacob realized that Teresa is a spy, but she isn't.
d. Ivan is younger than Penny, and Penny is younger than Ivan.
e. My sister is an only child.
f. Tom is an extraordinary violinist, but he isn't.

Contradictory fillers included logical contradictions (51a,d,e), presupposition failures (51b,c) and the speaker's negation of the statement introduced in the previous discourse with an unadorned declarative (51f).

Participants were asked to give each stimulus a score on a scale ranging from -3 (contradictory sentence) to 3 (non-contradictory sentence) with one point increments.

A hundred people were recruited via Amazon Mechanical Turk to participate in the study. All the participants self-identified as native speakers of English.

Задание:

- изучите результаты эксперимента в файле hw4q3.csv, подумайте, какими статистическими способами их можно было бы описывать.
- 1 балл: является ли различие рейтингов противоречивых филлеров (contradictory fillers, CF) и непротиворечивых филлеров (non-contradictory fillers, NCF) статистически значимым? Приведите код и его интерпретацию.
 - Придумайте, как извлечь информацию о рейтингах для каждой из переменных (CF, NCF в данном пункте)!
 - Чтобы считать данные из файла, используйте “,” в качестве разделителя, а не “;”:

```
>data=read.csv("C:/hw4q3.csv", header = TRUE, sep=",")
```

- 1 балл: является ли различие рейтингов непротиворечивых филлеров (NCF) и вложений (regular embeddings, RE) статистически значимым? Приведите код и его интерпретацию.
- 1 балл: является ли различие рейтингов вложений (RE) и сдвигов, расположенных внутри предложения, (sentence-internal shifts, IS) статистически значимым? Приведите код и его интерпретацию.
- 1 балл: является ли различие рейтингов вложений (RE) и сдвигов, расположенных в конце предложения, (sentence-final shifts, FS) статистически значимым? Приведите код и его интерпретацию.
- 1 балл: обоснуйте использование статистического теста. Подсказка: можно использовать Wilcoxon Rank Sum Test (<http://r-statistics.co/Statistical-Tests-in-R.html>).

4. *Финальный проект: начинаем! [1 балл, к занятию 15.12]*

- определиться с командой (в неё может входить от 1 до 3 человек)
- предварительно выбрать тему (список тем я отправлю в понедельник, 10.12)