

Критерий хи-квадрат Критерий Вилкоксона

Корпусные методы исследований языковых
процессов

Даша Попова

23.11.2022

<-

Alt + - (Windows)

Option + - (Mac)

Chi-Square test

Критерий хи-квадрат

- **Chi-Square test** is a statistical method to determine if two categorical variables have a significant correlation between them.
- Both those variables should be from the same population and they should be categorical like – Yes/No, Male/Female, Red/Green etc.
- This test only works for **categorical** data (data in categories), such as Gender {Men, Women} or color {Red, Yellow, Green, Blue} etc, but **not numerical** data such as height or weight.
- The numbers must be large enough. Preferably, each entry must be **5** or more.
- **The null hypothesis:** the variables are **independent**.
- The smaller the ***p*-value**, the higher the **significance** because it tells the investigator that the hypothesis under consideration may not adequately explain the observation.
- The null hypothesis H_0 is rejected if any of these probabilities is less than or equal to a small, fixed but arbitrarily pre-defined threshold value, which is referred to as the level of significance. Unlike the *p*-value, the α level is not derived from any observational data and does not depend on the underlying hypothesis; the value of α is instead set by the researcher before examining the data.
- The setting of α is arbitrary. By convention, α is commonly set to 0.05, 0.01, 0.005, or 0.001.
- The number of **degrees of freedom** (количество степеней свободы) is the number of values in the final calculation of a statistic that are free to vary.

<https://www.mathsisfun.com/data/chi-square-test.html>

		Верная гипотеза	
		H_0	H_1
Результат применения критерия	H_0	H_0 верно принята	H_0 неверно принята (Ошибка второго рода)
	H_1	H_0 неверно отвергнута (Ошибка первого рода)	H_0 верно отвергнута

Ошибка 1-го рода: нулевую гипотезу отвергают, когда она истинна, и делают вывод, что имеется эффект, когда в действительности его нет. Максимальный шанс (вероятность) допустить ошибку 1-го рода обозначается α (альфа). Это уровень значимости критерия; нулевую гипотезу отвергают, если наше значение p ниже уровня значимости, т. е., если $p < \alpha$.

Следует принять решение относительно значения α прежде, чем будут собраны данные; обычно назначают условное значение 0,05, хотя можно выбрать более ограничивающее значение, например 0,01.

Шанс допустить ошибку 1-го рода никогда не превысит выбранного уровня значимости, скажем $\alpha = 0,05$, так как нулевую гипотезу отвергают только тогда, когда $p < 0,05$. Если обнаружено, что $p > 0,05$, то нулевую гипотезу не отвергнут и, следовательно, не допустят ошибки 1-го рода.

<http://statistica.ru/theory/oshibki-pri-proverke-gipotez-moshchnost/>

Chi-Square test

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- O = the **Observed** (actual) value
- E = the **Expected** value

	Cat	Dog	
Men	207	282	489
Women	231	242	473
	438	524	962

	Cat	Dog	
Men	$\frac{(207-222,64)^2}{222,64}$	$\frac{(282-266,36)^2}{266,36}$	489
Women	$\frac{(231-215,36)^2}{215,36}$	$\frac{(242-257,64)^2}{257,64}$	473
	438	524	962

	Cat	Dog	
Men	$489 \times 438 / 962$	$489 \times 524 / 962$	489
Women	$473 \times 438 / 962$	$473 \times 524 / 962$	473
	438	524	962

	Cat	Dog	
Men	1,099	0,918	489
Women	1,136	0,949	473
	438	524	962

$$1,099 + 0,918 + 1,136 + 0,949 = 4,102$$

Chi-Square is 4,102

Multiply (rows - 1) by (columns - 1)

$$\text{Example: } DF = (2 - 1)(2 - 1) = 1 \times 1 = 1$$

$$p = 0,04283$$

Hypothesis: Gender and preference for cats or dogs are **independent**.

<https://www.mathsisfun.com/data/chi-square-test.html>

Chi-Square test

- `> dcmf <- matrix(c(207,282,231,242),ncol=2,byrow=TRUE)`
- `> colnames(dcmf) <- c("cat","dog")`
- `> rownames(dcmf) <- c("men","women")`
- `> dcmf <- as.table(dcmf)`
- `> dcmf`
- `> summary(dcmf)`

Number of cases in table: 962

Number of factors: 2

Test for independence of all factors: Chisq = 4.104, df = 1, p-value = 0.04279

- In this case, **p < 0,05**, so this result is significant: the variables are **not** independent.
- In other words, Gender is linked to Pet Preference: Men and Women have different preferences for Cats and Dogs.

Chi-Square test: plotting the table

- `> barplot(dcmf, legend=T, beside=T, main='Pet preferences by gender')`
- `> plot(dcmf, main="Pet preferences by gender")`
- `> mosaicplot(dcmf, main="Pet preferences by gender", xlab="Gender", ylab="Pet preference")`
- If you wish to switch which side (horizontal versus vertical) to determine the primary proportion then you can use the *sort* option. This can be used to switch whether the width or height is used for the first proportional length:
- `> mosaicplot(dcmf, sort=c(2,1))`

Chi-Square test

- For example, we can build a data set with observations on people's ice-cream buying pattern and try to correlate the gender of a person with the flavor of the ice-cream they prefer. If a correlation is found we can plan for appropriate stock of flavors by knowing the number of gender of people visiting.
- ```
> data=read.csv("C:/ice.csv", header = TRUE, sep=";")
```
- ИЛИ
- ```
> data <-  
read.csv("https://raw.githubusercontent.com/dashapopova/Corpus_methods_LangPolicy_2021/master/17.11/ice.csv",  
sep=";")
```
- ```
> pref.icecream = table(data$gender,data$flavour)
```
- ```
> pref.icecream
```
- **Способ 1:** -- лучше: если в клетках таблицы значения меньше 5, будет использоваться поправка Йейтса

```
> print(chisq.test(pref.icecream))
```

```
print(chisq.test(pref.icecream)$observed)
```

Считается, что критерий χ^2 не стоит применять, если хотя бы одно из **ожидаемых** значений меньше 5. Давайте посмотрим на ожидаемые наблюдения:

```
print(chisq.test(pref.icecream)$expected)
```

Если одно из **ожидаемых** значений меньше 5, то следует использовать тест Фишера: `fisher.test(pref.icecream)`

- **Способ 2:**
- ```
> summary(pref.icecream)
```

Зависит ли выбор мороженого от гендера?

**Look at the p-value:** If the p-value is less than 0.05, the gender of a person does correlate with the flavour of the ice-cream they prefer.

Нарисуйте пару графиков, которые адекватно отображали бы информацию о пристрастиях в выборе мороженого



# Критерий хи-квадрат

- Придумайте 10 задач, актуальных для языковой политики, где мог бы быть применён критерий хи-квадрат

# Wilcoxon Rank Sum Test

## Критерий Вилкоксона

- Wilcoxon rank test can be used to compare the medians of 2 samples
- Wilcoxon's rank sum test is a non-parametric test: it does not assume that the samples are driven from a normal distribution
- Use this test if the sample is small
- $H_0$ : медиана разницы в популяции равна нулю
- $H_1$ : медиана разницы в популяции не равна нулю
- $p\text{-value} < \alpha$ , we can reject the null hypothesis that both samples have same medians

# Wilcoxon Rank Sum Test

- `> x <- c(0.80, 0.83, 1.89, 1.04, 1.45, 1.38, 1.91, 1.64, 0.73, 1.46)`
- `> y <- c(1.15, 0.88, 0.90, 0.74, 1.21)`
- `> wilcox.test(x, y)`

# Wilcoxon Rank Sum Test

- `> x <- c(0.80, 0.83, 1.89, 1.04, 1.45, 1.38, 1.91, 1.64, 0.73, 1.46)`
- `> y <- c(1.15, 0.88, 0.90, 0.74, 1.21)`
- `> wilcox.test(x, y)`
- $H_0$ : медиана разницы в популяции равна нулю
- With a p-value of 0.25, we cannot reject the null hypothesis

# Wilcoxon Rank Sum Test

- Являются ли различия в рейтингах монолингвов и билингвов статистически значимыми?
- `> data=read.csv("C:/ice.csv", header = TRUE, sep=";")`
- `> summary(data)`
- `> wilcox.test(data$rating[data$nleng == "m"], data$rating[data$nleng == "b"])`
- Нет, не является, p-value = 0.117

# Wilcoxon Rank Sum Test

- Придумайте 10 задач, актуальных для языковой политики, где мог бы быть применён критерий Вилкоксона

# Практика - 1

## Пилотный эксперимент по изучению противоречий

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 slift (49b) or a sentence-final slift (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.

# Практика - 1

## Пилотный эксперимент по изучению противоречий

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).



# Практика - 1

## Пилотный эксперимент по изучению противоречий

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.

# Практика - 1

## Пилотный эксперимент по изучению противоречий

- ★ изучите результаты эксперимента в файле `contradictions.csv`, подумайте, какими статистическими способами их можно было бы описывать.
- ★ 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:/contradictions.csv", header = TRUE, sep=",")`
  - ◆ `data <-  
read.csv("https://raw.githubusercontent.com/dashapopova/Corpus_methods_LangPolicy_2021/master/17.11/contradictions.csv")`
- ★ является ли различие рейтингов непротиворечивых филлеров (NCF) и вложений (`regular embeddings`, RE) статистически значимым? Приведите код и его интерпретацию.
- ★ является ли различие рейтингов вложений (RE) и слифтов, расположенных внутри предложения, (`sentence-internal slifts`, IS) статистически значимым? Приведите код и его интерпретацию.
- ★ является ли различие рейтингов вложений (RE) и слифтов, расположенных в конце предложения, (`sentence-final slifts`, FS) статистически значимым? Приведите код и его интерпретацию.
- ★ проиллюстрируйте данные графиком, обоснуйте выбор графика

# Практика - 2

Из интервью с носителями одной деревни произвольным образом выбрали по пол часа и посчитали кол-во реализаций диалектных форм vs. недиалектных форм. В результате получилось, что у женщин было 107 диалектных форм vs. 93 недиалектных форм, а у мужчин — 74 vs. 45.

- ❖ Значима ли зафиксированная разница?
- ❖ Проиллюстрируйте данные графиком, обоснуйте выбор графика