

Школа лингвистики, 2024-25 уч. год**Теория вероятностей****Критерий χ^2 (29.04.2025)***И. В. Щуров, Д. А. Филимонов, Р. Я. Будылин*

Некоторые задачи основаны на книге Statistics, David Freedman, Robert Pisani, Roger Purves (fourth edition).

(a)		(b)		(c)		(d)	
значение	кол-во	значение	кол-во	значение	кол-во	значение	кол-во
1	5	1	9	1	90	1	10287
2	7	2	11	2	110	2	10056
3	17	3	10	3	100	3	9708
4	16	4	8	4	80	4	10080
5	8	5	12	5	120	5	9935
6	7	6	10	6	100	6	9934

Таблица 1: Результаты подбрасывания игральной кости

Задача 1. Результаты подбрасывания четырёх разных игральных костей показаны в таблице 1. С помощью критерия согласия Пирсона (критерия χ^2) проверить, является ли игральная кость «честной». Проверить нулевую гипотезу на уровнях значимости $\alpha = 1\%$ и $\alpha = 5\%$.

Задача 2. Для проверки авторства текста можно использовать часто встречающиеся слова (союзы, предлоги, некоторые характерные для конкретного корпуса текстов слова). Пусть было отобрано 5 слов и, по имеющемуся корпусу текстов, их частоты даны в первом столбце таблицы 2. В новом тексте количества встретившихся слов указаны во втором столбце таблицы 2. С помощью критерия χ^2 проверить, можно ли считать, что распределение этих слов такое же, как и в известном корпусе текстов автора. Проверить нулевую гипотезу на уровнях значимости $\alpha = 1\%$ и $\alpha = 5\%$.

Номер слова	Процент корпусу	Число в новом тексте
1	12	7
2	26	15
3	15	5
4	31	17
5	16	6

Таблица 2: Распределение частых слов

Задача 3. Пусть исследуется величина — число сложноподчинённых предложений из четырёх подряд идущих в тексте. Из большого корпуса выбраны случайные куски из 4 предложений и посчитано соответствующее количество. Можно ли считать это распределение биномиальным? Проверить нулевую гипотезу на уровнях значимости $\alpha = 10\%$ и $\alpha = 5\%$.

Число сложноподчинённых предложений	Количество в выборке
0	26
1	39
2	27
3	5
4	3

Таблица 3: Распределение сложноподчинённых предложений

Задача 4. (Гипотетический пример.) В некотором городе примерно миллион взрослых носителей языка. Выбирается случайная выборка из 10000 человек для изучения связи между полом и наличием дефекта речи при произнесении звука «р». Результаты приведены в таблице 4. С помощью теста χ^2 , проверить гипотезу о том, что пол и наличие дефекта речи независимы. Проверить нулевую гипотезу на уровнях значимости $\alpha = 1\%$ и $\alpha = 5\%$.

	Мужчин	Женщин
есть дефект	2792	3591
нет дефекта	1486	2131

Таблица 4: Результаты исследования

Задача 5. Демографы полагают, что в среднем рождается 55% мальчиков и 45% девочек. В некотором роддоме из 1000 случайно выбранных новорожденных оказалось 568 мальчиков. Применяя подходящий статистический критерий, проверить, согласуются ли эти данные с теорией. Проверить нулевую гипотезу на уровнях значимости $\alpha = 1\%$ и $\alpha = 5\%$.