

Лабораторная работа №4

Задание представлено в 6 вариантах. Для каждого вопроса требуется формализовать задачу и проверить статистическую гипотезу двумя критериями, если не сказано иное.

Вариант 1. В файле *exams_dataset.csv* (исходник [отсюда](#)) представлены данные об экзаменуемых.

1. Часто результаты интеллектуальных тестов аппроксимируют нормальным распределением, в частности, итоги IQ-тест можно приблизить нормальным распределением со средним 100 и стандартным отклонением 15. Можно ли приблизить результаты по математике нормальным распределением?
2. Можно ли утверждать, что результаты по чтению и письменной части принципиально не отличаются?
3. Есть подозрение, что посещавшие подготовительные курсы более успешны на экзаменах. Проверьте данное утверждение.

Вариант 2. В файле *mobile_phones.csv* (исходник [отсюда](#)) представлены данные о мобильных телефонах.

1. Разумно ли считать, что емкость аккумулятора распределена равномерно?
2. Верно ли, что телефонов с поддержкой 3G больше моделей с Wi-Fi? А разнится ли количество телефонов с touch screen от моделей с двумя сим-картами? На каждый вопрос по тесту.
3. Есть подозрение, что цена зависит от объема оперативной памяти. Проверьте данное утверждение.

Вариант 3. В файле *sex_bmi_smokers.csv* данные о пациентах.

1. Разумно ли индекс массы тела аппроксимировать нормальным законом?
2. Отличаются ли принципиально распределение индекса массы тела у мужчин и женщин?
3. Есть подозрение, что курящие склонны к ожирению. Кажется, что мужчины более склонны к ожирению. Проверьте данные утверждения, на каждую гипотезу по одному тесту.

Вариант 4. В файле *song_data.csv* (взято [отсюда](#)) приведены данные о музыкальных произведениях.

1. Разумно ли популярность песни аппроксимировать нормальным законом?
2. Отличается ли принципиально распределение рейтинга песни в зависимости от продолжительности (разбейте условно на "длинные" и "короткие", порог выбирайте сами)?

3. Зависит ли популярность песни от продолжительности?

Вариант 5. В файле *MEN_SHOES.csv* (источник [отсюда](#)) приведены данные о продажах мужской обуви.

1. Разумно ли количество проданных экземпляров обуви аппроксимировать распределением Пуассона, а рейтинг – нормальным распределением (по 1 тесту на каждый вопрос)?
2. Верно ли что распределения количества проданных экземпляров существенно не отличаются в зависимости от бренда? Тот же вопрос для цены (по одному тесту на утверждение).
3. Есть подозрение, что рейтинг зависит от цены. Проверить данное предположение.

Вариант 6. В файле *cars93.csv* приведены данные об авто.

1. Разумно ли мощность считать равномерно распределенной, а цену – нормально (для каждого теста по вопросу)?
2. Верно ли, что распределения мощности для каждого типа авто принципиально не отличаются? Тот же вопрос про цену (для каждого вопроса по тесту).
3. Есть подозрение, что цена зависит от мощности авто. Проверьте данное предположение.

Ключевые понятия:

- Постановка задачи проверки статистических гипотез.
- Статистический критерий и его статистика. Области принятия и опровержения нулевой гипотезы, p-value.
- Ошибки I и II рода.
- Критерии согласия. Примеры критериев.
- Критерии однородности. Примеры критериев.
- Критерии независимости. Примеры критериев.