

Практика 8. Сравнение МО двух нормальных случайных величин Достоверность показаний зашумленных измерений

Практика является продолжением практики о зашумленных измерениях.

Часть 1. Функции проверки гипотез в случае известных дисперсий

1. Реализовать функцию расчета статистики e и z по двум выборкам $P1$, $P2$:

```
function e = e_statistic(P1, P2)
function z = z_statistic(P1, P2, D1, D2)
```

Результатом обеих функций являются скалярные величины e , z , учитывающие усреднение по выборке. Формулы расчета обеих статистик должны в коде фигурировать ровно один раз в этих функциях. Глобальные переменные в данных функциях запрещены. В дальнейшем при расчете статистик e , z обязательно вызывать эти функции. Код с нарушением этого требования не принимается.

2. Осознать, почему на входе функции расчета e -статистики не нужны дисперсии случайных составляющих погрешности, а z -статистики нужны. Осознать, почему не передается объем выборки n .
3. Реализовать функцию проверки гипотезы о равенстве МО двух СВ на основе статистик e и z .

```
function [h, border_low, border_high] = test_equal_means_e(e, D1, D2, n, alpha)
function [h, border_low, border_high] = test_equal_means_z(z, alpha)
```

h – результат проверки гипотезы

$border_low$, $border_high$ – границы области принятия основной гипотезы по e или z статистикам.

Далее везде в коде для проверки гипотез использовать только эти функции. Код с нарушением этого требования не принимается.

4. Осознать, почему в функции расчета e -статистики не нужны дисперсии случайных составляющих погрешности, а для z -статистики нужны.

Часть 2. Функции проверки гипотез в случае неизвестных дисперсий

1. В дополнение к функциям $e_statistic$, $z_statistic$ реализовать функцию расчета t -статистики e и z по двум выборкам $P1$, $P2$:

```
function t = t_statistic(P1, P2)
```

Результатом функции является скалярная величина t , учитывающая усреднение по выборке. Формулы расчета обеих статистик должны в коде фигурировать ровно один раз в этих функциях. Глобальные переменные в данных функциях запрещены. В дальнейшем при расчете статистик e , z обязательно вызывать эти функции. Код с нарушением этого требования не принимается.

2. Осознать, почему на входе функции расчета e -статистики не нужны дисперсии случайных составляющих погрешности, а z -статистики нужны. Осознать, почему не передается объем выборки n .

3. В дополнение к функциям $test_equal_means_e$, $test_equal_means_z$ реализовать функцию проверки гипотезы о равенстве МО двух СВ на основе статистики t .

```
function [h, border_low, border_high] = test_equal_means_t(t, n, alpha)
```

n – объем выборки, по которому посчитана статистика t ,

h – результат проверки гипотезы,

$border_low$, $border_high$ – границы области принятия основной гипотезы по t -статистике.

Далее везде в коде для проверки гипотез использовать только эту функцию. Код с нарушением этого требования не принимается.

4. Осознать, чем отличается t -статистика от z -статистики.

Часть 3. Распределения статистик в условиях H_0 и H_1

1. Задаться объемом измерительной выборки $5 < n < 20$.
2. Для статистики e на одном графике построить:
 - a. теоретическую плотность распределения статистики Критерия в условиях основной гипотезы (показания датчиков «в среднем» равны, отличия обусловлены случайной составляющей погрешности и носят чисто случайный характер);
 - b. теоретическая плотность распределения статистики Критерия в условиях альтернативной гипотезы;
 - c. в тех же условиях методом статистического моделирования Монте-Карло ($N \geq 1000$) построить полигон относительных частот статистики Критерия (hist_density);
 - d. методом статистического моделирования ($N \geq 1000$) построить полигон относительных частот статистики Критерия при условии альтернативной гипотезы;
 - e. осознать, чем отличается n от N .
3. Построить то же самое для статистики z .
4. Построить то же самое для статистики t .
5. На построенные в этой части задания графики (их должно быть три) нанести нижнее и верхнее пороговые значения для принятия или отклонения основной гипотезы. Пороги брать из функций test_equal_means_e, test_equal_means_z, test_equal_means_t.
6. Подобрать n , на котором визуально видна разница между t -статистикой и z -статистикой. В чем она заключается?

Часть 4. Исследование статистических свойств критериев e , z

1. Задаться объемом измерительной выборки $5 < n < 20$.
2. Методом Монте-Карло оценить ошибку **первого** рода при принятии гипотезы о равенстве MO нормальных СВ на основе e -статистики.
Для этого сгенерировать $N \geq 1000$ реализаций e -статистики в условиях основной гипотезы и посчитать долю случаев, в которых функция test_equal_means_e принимает альтернативную гипотезу.
3. Аналогичным образом оценить ошибку **второго** рода при принятии гипотезы о равенстве MO нормальных СВ на основе e -статистики
4. Посчитать ошибки первого и второго рода для z -статистики.
5. Увеличить объем измерительной выборки n в два раза. Проверить ошибки первого и второго рода. Обосновать их изменение или, наоборот, неизменность.

Часть 5. Исследование ошибок первого и второго рода для статистик t и z

1. Методом статистического моделирования исследовать ошибки первого и второго рода в зависимости от объема выборки для z -статистики и t -статистики.
Для этого построить графики ошибок 1-го, 2-го рода в зависимости от объема выборки n для z -статистики и t -статистики. Все 4 графика построить на одном, объяснить, чем вызвано их отличие.
2. Методом статистического моделирования исследовать ошибку первого и второго рода в зависимости от близости систематических погрешностей датчиков.
Для этого построить графики ошибок 1-го, 2-го рода в зависимости от разницы систематических погрешностей для z -статистики и t -статистики. Все 4 графика построить на одном, объяснить, чем вызвано их отличие.
3. Задаться требуемыми ошибками первого и второго рода (не обязательно одинаковыми), определить по построенным графикам
 - a. объемы выборки, при котором эти ошибки будут обеспечены.
 - b. разницу систематических погрешностей, которую можно выявить с такими ошибками

Вопросы к защите

1. Вероятностная модель измерения двумя датчиками одного физического параметра.
 - a. Записать вероятностную модель генератора сигнала датчика «до измерений» с учетом истинного значения параметра, аналитически вывести его МО, дисперсию. Вывести МО, дисперсию для разницы показаний датчиков.
 - b. Эскиз графика матожидания показаний датчика от времени.
 - c. Определение случайного процесса. Почему сигналы датчиков являются случайными процессами? Определение стационарного случайного процесса. Являются ли они стационарными? Те же вопросы для разницы показаний датчиков.
 - d. Что такое случайная и систематическая погрешности? Как соотнести термины случайная и систематическая погрешности с параметрами распределения показаний случайной величины?
2. Распределения статистик критерия
 - a. Формула расчета усредненной статистики \bar{e}_n по двум временным рядам показаний датчиков
 - b. Вывести распределения статистики \bar{e}_n в условиях основной и альтернативной гипотезы. Как связана дисперсия средневывборочного с дисперсией исходной СВ?
 - c. Вывести распределения статистики z_n в условиях основной и альтернативной гипотезы.
 - d. Как надо модифицировать расчет статистики e , чтобы учесть изменение физического параметра? Почему нельзя использовать расчет по учебникам?
3. Проверка гипотез
 - a. Что такое статистика критерия? Какие статистики использовались в данном ДЗ?
 - b. Что такое критическая область и область принятия гипотезы?
 - c. Почему в функции расчета e -статистики не нужны дисперсии случайных составляющих погрешности, а для z -статистики нужны?
 - d. Что такое ошибка первого и второго рода? Как посчитать ошибку первого и второго рода для e и z -статистик?
 - e. Что такое простая и сложная гипотеза? Какая из двух гипотез рассматриваемых критерием является простой, а какая сложной?
 - f. Как связаны плотность распределения вероятности и функция распределения вероятности. Что такое инверсная функция распределения вероятностей? Как с ее помощью посчитать пороги для e ?
 - g. Как влияет объем выборки на ошибку первого рода?
 - h. Как влияет объем выборки на ошибку второго рода?
4. Вопросы по t -статистике
 - a. Формула расчета t -статистики
 - b. Почему расчет t -статистики корректен, несмотря на нестационарность измерительных временных рядов $P1, P2$?
 - c. Сформулировать основную и альтернативную гипотезы, которые проверялись с помощью t -статистики.
 - d. Что будет, если пороги для t -статистики считать так, как будто, это z -статистика?
 - e. Какое распределение у t -статистики в условиях основной и альтернативной гипотезы? Какие распределения у z -статистики в этих же условиях. Нарисовать эскиз, демонстрирующий отличия между этими статистиками?
 - f. Какие факторы влияют на ошибки первого и второго рода? Чтобы подготовить ответ, надо вспомнить эксперименты с e, z, t -статистиками из двух ДЗ.
 - g. Можно ли воспользоваться t -статистикой, если известны дисперсии случайных составляющих погрешности?

- h. Что лучше: z-статистика или t-статистика?
- i. Как выбрать объем выборки, учитывая, что с технической точки зрения лучше ее взять меньше, чтобы уменьшить время сбора показаний датчиков для анализа.
- j. В каких случаях нельзя использовать z-статистику, но можно t-статистику?
- k. В каких случаях нельзя использовать z-статистику и t-статистику?