

## Практика 4. Сравнение МО двух нормальных случайных величин в случае НЕизвестных дисперсий.

### Достоверность показаний датчиков с НЕизвестными случайными погрешностями

Задание рекомендуется выполнить на основе аналогичного задания с известными дисперсиями. Нужно взять за основу код прошлого задания, но обязательно сохранить его версию с известными дисперсиями.

Основные доработки коснутся части 4.

#### Часть 1. Подготовка генераторов данных

В задании с известными дисперсиями было принято, что истинное значение параметра не изменяется во времени. Однако это очень сильное упрощение, поскольку в реальной жизни параметр изменяется, и датчик собственно и нужен для того, чтобы фиксировать это изменение. Более правдоподобна следующая вероятностная модель показаний датчиков:

$$\begin{aligned}\hat{P}_1(t) &= P_{\text{ист}}(t) + \Delta_1 + \delta_1 \\ \hat{P}_2(t) &= P_{\text{ист}}(t) + \Delta_2 + \delta_2\end{aligned}$$

Обратите внимание, что теперь матожидание измерений зависит от времени:

$$\begin{aligned}E\hat{P}_1(t) &= P_{\text{ист}}(t) + \Delta_1 \\ E\hat{P}_2(t) &= P_{\text{ист}}(t) + \Delta_2\end{aligned}$$

Учтите этот факт в дальнейшем при расчете t-статистики.

Модифицируйте генератор из задания с известными дисперсиями, чтобы учесть изменение истинного значения параметра. Закон изменения  $P_{\text{ист}}(t)$  взять аperiodический, либо периодический.

#### Часть 2. Функции проверки гипотез

1. В дополнение к функциям `e_statistic`, `z_statistic` реализовать функцию расчета t-статистики `e` и `z` по двум выборкам `x1`, `x2`:

```
function t = t_statistic(x1, x2)
```

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, alpha)
```

`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

1. Строятся графики, описанные в части 3 задания с известными дисперсиями для z-статистики и t-статистики. График для e-статистики строить не нужно.
2. Подобрать n, на котором визуально видна разница между t-статистикой и z-статистикой. В чем она заключается?

#### **Часть 4. Исследование ошибок первого и второго рода**

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.
2. Методом статистического моделирования исследовать ошибку 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.
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

#### **Часть 5. Спроектировать систему проверки достоверности показаний датчиков**

Вопрос тот, же, что и в ДЗ с известными дисперсиями. Какие новые мысли по нему появились у Вас при выполнении данного ДЗ?

Как выбрать объем выборки для анализа?

#### **Вопросы к защите**

1. Каким образом следует модифицировать расчет t-статистики, чтобы расчет был корректен при изменяющемся во времени  $P_{ист}(t)$ .
2. Сформулировать основную и альтернативную гипотезы, которые проверялись с помощью t-статистики.
3. 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?
4. Какие факторы влияют на ошибки первого и второго рода? Чтобы подготовить ответ, надо вспомнить эксперименты с e, z, t-статистиками из двух ДЗ.
5. Можно ли воспользоваться t-статистикой, если известны дисперсии случайных составляющих погрешности?
6. Что лучше: z-статистика или t-статистика?
7. Как выбрать объем выборки, учитывая, что с технической точки зрения лучше ее взять меньше, чтобы уменьшить время сбора показаний датчиков для анализа.
8. В каких случаях нельзя использовать z-статистику, но можно t-статистику?
9. В каких случаях нельзя использовать z-статистику и t-статистику?