

Образовательный портал “GeekBrains”  
факультет “Data Science в медицине”  
курс “теория вероятности и математическая статистика”

## **ИТОГОВЫЙ ПРОЕКТ**

“Автоматизация расчета концентрации иммуноглобулинов класса G к вирусу кори в сыворотке крови”

выполнила: Гончарова А.В.

## **СОДЕРЖАНИЕ**

1. Введение
2. Цель и задачи
3. Статистические методы
4. Получение данных
5. Статистическая обработка полученных данных
6. Реализация цели программно на python

## **ВВЕДЕНИЕ**

Метод определения концентрации иммуноглобулинов класса G к вирусу кори в сыворотке крови основан на иммуноферментном анализе (ИФА) с применением рекомбинантного антигена вируса кори.

При ИФА получают растворы разной степени окрашивания, которую оценивают по величине оптической плотности (ОП).

Степень окраски пропорциональна концентрации иммуноглобулина G (IgG) к вирусу кори в анализируемых образцах.

В набор входят реактивы для постановки ИФА, 6 (шесть) стандартных растворов (калибраторов) с известной концентрацией IgG к вирусу кори (значения предоставляются производителем) и 96-луночный планшет.

В ходе постановки ИФА исследуемые образцы (калибраторы и сыворотки крови пациентов) добавляют в лунки планшета (1 лунка = 1 образец) и последовательно вносят в каждую лунку реагенты согласно схеме проведения иммуноферментного анализа.

В итоге ифа-планшет помещают в спектрофотометр и получают оптическую плотность реакционной смеси каждой лунки.

После измерения величины оптической плотности растворов на основании калибровочного графика рассчитывается концентрация IgG к вирусу кори.

Калибровочный график строят для каждой постановки ИФА, вручную, что может приводить к погрешности построения и, следовательно, расчета концентрации.

## ЦЕЛЬ И ЗАДАЧИ

**Цель работы:** автоматизировать построение калибровочного графика и расчет концентрации IgG к вирусу кори.

**Задачи:**

1. Получить значения оптической плотности для калибраторов
2. Оценить силу и направленность связи между ОП и концентрацией IgG к вирусу кори
3. Рассчитать коэффициенты модели линейной регрессии
4. Оценить статистическую значимость модели и коэффициентов
5. Построить калибровочный график
6. С помощью построенной регрессионной модели рассчитать концентрацию IgG к вирусу кори в исследуемых образцах

## СТАТИСТИЧЕСКИЕ МЕТОДЫ

В работе использовали следующие методы математической статистики:

**Среднее арифметическое:** среднее арифметическое - это сумма всех наблюдений, деленная на количество наблюдений ( $n$ ).

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

**Корреляционный анализ:** корреляционный анализ - определение взаимосвязи между переменными. Наличие связи, её направленность и силу оценивают по коэффициенту корреляции ( $r$ ).

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Коэффициент корреляции принимает значения от  $-1$  до  $+1$ .

**направленность:**

от  $0$  до  $1$  - связь прямая (переменная увеличивается при увеличении второй переменной)

от  $-1$  до  $0$  - связь обратная (одна переменная уменьшается с увеличением второй переменной)

### ***сила связи:***

$|r|$  от 0 до 0,3 – слабая связь

$|r|$  от 0,3 до 0,7 – умеренная связь

$|r|$  от 0,7 до 1,0 – сильная связь

**Линейный регрессионный анализ:** регрессионный анализ — это статистический метод, используемый для описания взаимосвязи между двумя переменными и для прогнозирования одной переменной по значению другой.

Уравнение для линейной регрессии имеет вид:

$$y = a + b \cdot x$$

где

$y$  – прогнозируемая переменная (таргетная)

$x$  – предсказывающая переменная (предиктор)

$a, b$  – коэффициенты регрессии

$$b = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - (\bar{x})^2};$$

$$a = \bar{y} - b \cdot \bar{x}.$$

### **Оценка регрессионной модели:**

**коэффициент детерминации  $R^2$ :** это доля изменения зависимой переменной, объясняемая регрессионной моделью, и является мерой качества соответствия модели. Он может принимать значения от 0 до 1 и рассчитывается следующим образом:

$$R^2 = r_{xy}^2$$

**Ф-критерий Фишера:** F-критерий Фишера позволяет оценить значимость уравнения линейной регрессии в целом. Фактическую величину F-критерия для парной линейной регрессии (когда есть только один факторный признак) можно вычислить по формуле, включающей коэффициент детерминации:

$$F_{\text{факт}} = \frac{r_{xy}^2 \cdot (n - 2)}{(1 - r_{xy}^2)}$$

где  $n$  - это число наблюдений.

Критическое значение F-критерия находят по таблице.

Если  $F_{\text{факт}} > F_{\text{кр}}$ , то уравнение регрессии признается статистически значимым.

**t-критерий Стьюдента:** оценка статистической значимости параметров регрессии проводится с помощью t-статистики Стьюдента.

Наблюдаемые значения критерия находят по формулам:

$$t_a = \frac{a}{m_a} \text{ и } t_b = \frac{b}{m_b}$$

$$m_a = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{n \cdot \sigma_x}$$

$$m_b = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}}$$

Если расчетное значение критерия  $t_a, t_b$  больше табличного значения  $t_{\text{кр}}$  при заданном уровне значимости  $\alpha$  (0,1; 0,05; 0,01), то коэффициент регрессии считается значимым.

Чтобы найти табличное значение t-статистики Стьюдента, нужно знать число степеней свободы и уровень значимости  $\alpha$ . Число степеней свободы равно  $df = n - m - 1$ , где  $n$  - число наблюдений и  $m$  - число признаков (факторов). Для парной регрессии число степеней свободы будет равно  $n - 2$ .

## ПОЛУЧЕНИЕ ДАННЫХ

Выполнена постановка иммуноферментного анализа (ИФА).

В итоге получили в каждой лунке планшета реакционные смеси разной степени окрашивания.

Загрузили планшет в спектрофотометр и считали значения оптической плотности (ОП) для каждой лунки.

В таблице 1 представлены результаты ОП для калибраторов и исследуемых образцов сыворотки крови.

Таблица 1.

образец	оптическая плотность	
	измерение 1	измерение 2
калибратор 1	0,167	0,157
калибратор 2	0,279	0,275
калибратор 3	0,419	0,402
калибратор 4	0,671	0,638
калибратор 5	1,165	1,089
калибратор 6	2,144	1,995
пациент 1	1,505	1,515
пациент 2	0,998	0,924
пациент 3	0,615	0,688

## СТАТИСТИЧЕСКАЯ ОБРАБОТКА ПОЛУЧЕННЫХ ДАННЫХ

1. Расчет среднего арифметического ОП по двум измерениям (дублям) для калибраторов и образцов (Табл.2).

Таблица 2.

образец	оптическая плотность
калибратор 1	0,162
калибратор 2	0,277
калибратор 3	0,4105
калибратор 4	0,6545
калибратор 5	1,127
калибратор 6	2,0695
пациент 1	1,51
пациент 2	0,961
пациент 3	0,6515

2. Определение связи между ОП и концентрации IgG (табл. 3).

Таблица 3.

образец	ОП	концентрация*
калибратор 1	0,162	1
калибратор 2	0,277	1,93
калибратор 3	0,4105	6,63
калибратор 4	0,6545	10,23
калибратор 5	1,127	12,6
калибратор 6	2,0695	23,7

\* - значения предоставляются производителем набора

Расчет коэффициента корреляции:

$$r = 0,9816$$

Таким образом, мы подтвердили прямую сильную корреляционную связь между оптической плотностью и концентрацией IgG к вирусу кори.

3. Построение регрессионной модели расчета концентрации по значению оптической плотности.

Уравнение имеет вид:  $y = a + b * x$

$$a = 0,38$$

$$b = 11,45$$

Получили *уравнение модели линейной регрессии*:

$$\text{Концентрация IgG} = 0,38 + 11,45 * \text{ОП}$$

**Коэффициент детерминации:**  $R^2 = 0,9636$

Таким образом, 96,36% вариации концентрации IgG (y) объясняется вариацией фактора x – оптическая плотность.

**F-критерий Фишера для модели:**

По таблице находим критическое значение для F ( $\alpha = 0.05$ )

$$F_{\text{кр}} = 7,71$$

Рассчитываем фактическое значение для F:

$$F_{\text{факт}} = 105,92$$

Так как  $F_{\text{факт}} = 105,92 > F_{\text{кр}} = 7,71$ , то уравнение регрессии статистически значимо.

**t-статистика Стьюдента:**

По таблице находим критическое значение для t ( $\alpha = 0.05$ )

$$t_{\text{крит}} = 2,776$$

Рассчитываем фактическое значение для  $t_a$  и  $t_b$ :

$$t_a = 0,34$$

$$t_b = 10,29$$

t-статистика Стьюдента для коэффициента регрессии  $a < t_{\text{крит}}$ , то есть статистически не значим.

t-статистика Стьюдента для коэффициента регрессии  $b > t_{\text{крит}}$ , то есть статистически значим.

### **Заключение по построенной модели линейной регрессии:**

Построенная модель линейной регрессии для расчета концентрации IgG к вирусу кори признана статистически значимой и может использоваться в практической работе.

4. Расчет концентрации IgG к вирусу кори в образцах пациентов с помощью построенной модели линейной регрессии.

Таблица 4.

образец	ОП	концентрация* $0,38+11,45*ОП$
пациент 1	1,51	17,66
пациент 2	0,961	11,38
пациент 3	0,6515	7,84

### **РЕАЛИЗАЦИЯ РАБОТЫ НА PYTHON**

Программный код представлен в виде файла jupyter notebook (GoncharovaAV\_finall\_project.ipynb).