

# Лабораторная работа №7

## Анализ выборки. Критерий согласия Пирсона.

К.С. Пилипенко 

2023

Критерий согласия Пирсона (Хи-квадрат) был придуман для проверки значимости расхождения эмпирических (наблюдаемых) и теоретических (ожидаемых) частот. Выражается следующей формулой:

$$\chi^2 = \sum_i^n \frac{(O_i - E_i)^2}{E_i}, \quad (1)$$

где  $O_i$  — наблюдаемые частоты (Observed),  $E_i$  — ожидаемые частоты (Expected).

Полученное значение  $\chi^2$  сравнивают с теоретически рассчитанным критическим значением  $\chi_{кр.}^2$ , которое зависит от значения доверительной вероятности (как правило принимается равным 95%) и числа степеней свободы  $k$ , которое на один меньше количества уникальных значений в выборке ( $k = N - 1$ ).

Для расчета критического значения критерия  $\chi_{кр.}^2$  можно воспользоваться специальной таблицей, но лучше и проще всего воспользоваться функцией `ХИ2.ОБР.ПХ`

Для получения нормального распределения используется функция `НОРМ.РАСП( $x$ ;  $\bar{x}$ ;  $\sigma$ ; интегральная)`, где

$x$  — значение, для которого строится распределение.

$\bar{x}$  — среднее арифметическое распределения.

$\sigma$  — стандартное отклонение распределения.

интегральная — логическое значение, определяющее форму функции. Если ИСТИНА функция возвращает значение функции распределения  $f(x)$ , если ЛОЖЬ то, возвращается значение функции плотности вероятности  $\omega(x)$ .

Последний аргумент является необязательным и по умолчанию стоит как ЛОЖЬ. Более подробно о синтаксисе этой функции можно узнать [здесь](#).

## Ход работы

### Задание №1. Генерация выборки

- Создайте файл с расширением `.xlsm`. По умолчанию вкладка Разработчик скрыта, поэтому надо её включить. Заходим по пути Файл→Параметры→Настроить ленту и ставим галочку на вкладке Разработчик;
- В разделе Разработчик нажмите на макросы, дайте любое название макроса и создайте его. Далее в открывшемся окне введите код из листинга 1;

```
1      Sub NormDistGen ()  
2          ' Сюда надо поместить код из листинга  
3      End Sub
```

- При запуске кода в первом столбце таблицы появится сгенерированная выборка объемом 500 элементов;

### Задание №2. Получение наблюдаемых и ожидаемых частот

- Найдём максимальный и минимальный элементы выборки с помощью функций МАКС и МИН соответственно.
- Создайте столбец из целых чисел, значения которого будут использованы в качестве границ интервалов ( $N$ ). Первое и последнее числа столбца должны быть больше минимального и максимального значения выборки соответственно.

*Пример. Если  $x_{min} = -0,961$  и  $x_{max} = 25,983$  то получится столбец от 0 до 26 с шагом 1;*

- Следующий шаг — получить частоты попадания элементов выборки в эти интервалы. Для этого необходимо воспользоваться функцией `ЧАСТОТА(массив данных; массив верхних границ диапазонов)`. Перед тем как задавать функцию нужно выделить диапазон ячеек, размер которого соответствует размеру массива границ. После нажатия комбинации `CTRL+SHIFT+ENTER` сформируется массив частот  $O(N)$ ;
- Постройте график зависимости массива частот  $O$  от массива границ  $N$ ;
- Теперь нужно получить столбец  $E(N)$ . Перед этим нужно посчитать столбец вероятностей  $(P_i)$  для каждого  $N$  (значения из массива границ) с помощью функции `НОРМ.РАСП` с пустым последним параметром. Остаётся открытым вопрос, откуда брать среднее  $(\bar{N})$  и  $\sigma$ ? Эти параметры можно подобрать вручную построив два графика в одних координатах  $E(N)$  и  $O(N)$ , и добившись их наилучшего соответствия;
- Чтобы найти ожидаемые (теоретические) частоты  $E_i$  нужно умножить соответствующие вероятности  $(P_i)$  на объём выборки, то есть на 500. Ож. частоты  $E_i$  должны быть целыми ненулевыми числами, поэтому стоит использовать функцию `ОКРУГЛВВЕРХ`;

### **Задание №3. Оценка критерия Пирсона**

- Используя формулу 1 посчитать критерий Пирсона;
- Используя функцию `ХИ2.РАСП` постройте график плотности распределения  $\omega_{\chi^2_{кр.}}(N)$  (последний параметр должен быть ЛОЖЬ);
- В отдельном столбце посчитать критическое значение пользуясь функцией `ХИ2.ОБР.ПХ`. Указать это значение на графике  $\omega(\chi^2_{кр.})$ . Сравнить полученное значение с экспериментальным;
- Получить p-value используя функцию `ХИ2.РАСП.ПХ` для посчитанного  $\chi^2$
- Чтобы убедиться в правильности полученного результата сравните p-value со значением функции `ХИ2.ТЕСТ`, которая принимает на вход массивы наблюдаемых и ожидаемых частот.

## Контрольные вопросы

1. Сформулируйте нулевую гипотезу  $H_0$ . Назовите условие, при котором можно отклонить нулевую гипотезу.
2. Можно ли в программе Excel сгенерировать выборку имеющую нормальное распределение? Если можно, то как это реализовать?
3. Что такое p-value (p-значение)? Чему численно равно это значение?
4. В каких случаях используется функция ХИ2.ОБР? Что она позволяет оценить

## Приложение

Листинг 1: Код генератора выборки с нормальным распределением

```
1      Randomize
2      Dim i As Long
3      Dim mean As Integer
4      Dim sigma As Integer
5      Dim random As Double
6
7      mean = 10 * Rnd + 6
8      sigma = 3 * Rnd + 2
9      i = 501
10     Range("A1").Select
11     For i = 2 To i
12         random = WorksheetFunction.NormInv(Rnd, mean,
13             sigma)
14         ActiveCell.Value = random
15         ActiveCell.Offset(1, 0).Select
16     Next i
```