

В исторических исследованиях часто возникают задачи анализа какого-либо признака, свойственного большой группе объектов. В таблице приведены данные по наделам земли в десятинах на душу населения для различных губерний царской России.

N	Губерния	Посев на душу (дес.)	N	Губерния	Посев на душу (дес.)
1.	Архангельская	0,19	26.	Новгородская	0,48
2.	Астраханская	0,38	27.	Олонецкая	0,37
3.	Бессарабская	0,99	28.	Оренбургская	0,94
4.	Виленская	0,67	29.	Орловская	0,81
5.	Витебская	0,6	30.	Пензенская	0,94
6.	Владимирская	0,57	31.	Пермская	0,64
7.	Вологодская	0,45	32.	Петербургская	0,42
8.	Волынская	0,57	33.	Подольская	0,66
9.	Воронежская	0,89	34.	Полтавская	0,73
10.	Вятская	0,99	35.	Псковская	0,54
11.	Гродненская	0,59	36.	Рязанская	0,75
12.	Донская	1,55	37.	Самарская	1,14
13.	Екатеринославская	1,26	38.	Саратовская	1,06
14.	Казанская	0,86	39.	Симбирская	0,91
15.	Калужская	0,5	40.	Смоленская	0,58
16.	Киевская	0,52	41.	Таврическая	1,95
17.	Ковенская	0,61	42.	Тамбовская	0,88
18.	Костромская	0,64	43.	Тверская	0,5
19.	Курляндская	0,76	44.	Тульская	0,9
20.	Курская	0,87	45.	Уфимская	0,71
21.	Лифляндская	0,66	46.	Харьковская	0,9
22.	Минская	0,6	47.	Херсонская	1,62
23.	Могилевская	0,56	48.	Черниговская	0,69

24.	Московская	0,35	49.	Эстляндская	0,56
25.	Нижегородская	0,65	50.	Ярославская	0,52

Рис.1 Таблица "Посев на душу населения"

Здесь через N обозначены порядковые номера губерний, расположенных в алфавитном порядке. Через x обозначена величина посевного надела земли в десятинах. Эту величину x можно принять в качестве статистического признака, характеризующего конкретную губернию, а значение этого признака может объяснять те или иные социально-исторические процессы, имевшие место в соответствующих губерниях. Методы математической статистики позволяют сгруппировать губернии по значению указанного признака.

Для этого прежде всего необходимо расположить губернии в строках таблицы в порядке возрастания величины x . Эта операция называется ранжированием по данному признаку. Потом необходимо весь диапазон изменения величины x поделить на равные интервалы, середины которых мы будем обозначать через x_i , и определить число губерний n_i , попадающих в каждый интервал x_i . Далее надо определить частоту f_i следующим образом $f_i = n_i/n$, где n называют объемом выборки.

Изменение величины f_i в зависимости от x_i называют рядом распределения статистической величины, а график функции $f_i(x_i)$ называют гистограммой. Функция $f_i(x_i)$ обладает некоторыми статистическими характеристиками, которые также проливают свет на некоторые социально-исторические явления.

Стандартный набор статистических характеристик обычно включает в себя среднее арифметическое значение, дисперсию, среднее квадратичное отклонение, а также минимальное и максимальное значения признака.

Средним арифметическим значением признака x называется величина

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1),$$

где x_i - значение признака для i - го объекта, n - общее количество объектов. Среднее арифметическое является характеристикой расположения признака x и представляет собой типичное значение признака, центральную точку его распределения.

Дисперсия $D(x)$ есть среднее арифметическое квадратов отклонений значений признака от среднего арифметического, т.е.:

$$D(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (2)$$

При небольших объемах выборки используется другая формула, позволяющая получить несмещенную оценку:

$$D(x) = \frac{n \sum_{i=1}^n x_i^2 - (\sum_{j=1}^n x_j)^2}{n(n-1)}. \quad (3)$$

Дисперсия является параметром, показывающим, насколько широк разброс значений признака по каждую сторону от типичного значения. Параметр такого рода называется характеристикой рассеяния или характеристикой сосредоточения. Рассеяние и сосредоточение находятся, конечно же, в обратной зависимости: чем больше рассеяние, тем меньше сосредоточение, и наоборот.

Средним квадратическим отклонением (стандартным отклонением) признака x называется величина:

$$\delta(x) = \sqrt{D(x)}. \quad (4)$$

Стандартное отклонение является более удобной характеристикой, чем дисперсия, так как измеряется в тех же единицах, что и сам признак.

	A	B	C	D	E	F	G	H
1	№	Губерния	посев на душу	Массив интервалов (x_i)	Частота попад. в инт. (n_i)	Частота $f_i = n_i/n$	Параметры оценивания для $n=50$	
2	1	Архангельская	0,19				Среднее \bar{x}	
3	2	Астраханская	0,38				Дисперсия $D(x)$	
4	3	Бессарабская	0,99				Станд.откл. $\delta(x)$	
5	4	Виленская	0,67					

Рис.2 Таблица "Посев на душу населения"

Технология решения задачи:

1. В ячейки A1:F1 (рис.2) занесите заголовки столбцов. Перенос на другую строку в ячейке осуществляется с помощью нажатия клавиши Alt + Enter.
2. В ячейку A2 запишите начальный номер 1.
3. Выполните операцию автозаполнения столбца A значениями N.
4. В ячейки B2:B51 запишите названия губерний.

5. В ячейки C2:C51 занесите значения посевов на душу населения (величина x).
6. Выполните ранжирование по значению величины x . Для этого выполните следующие действия:
 - выделите столбцы A, B, C;
 - выполните команду: *Данные / Сортировка*;
 - выберите: *Сортировать по Посев на душу*.

В итоге получится список губерний, ранжированный по величине Посев на душу.

7. Осуществите группировку данных по интервалам. Для этого в столбец D необходимо ввести массив интервалов. Минимальное число массива должно быть меньше минимального значения признака, равного 0,19. Выберите его равным нулю и введите в ячейку D2. Заполните ячейки D2:D32 рядом значений арифметической прогрессии от 0 до 3 с шагом 0,1.
8. Определите частоты попадания результатов в каждую группу. Для этого выделите ячейки E2:E32, щелкните по кнопке «Мастер функций». В графе *Категория* выберите *Статистические*. В графе *Функция* выберите *Частота*. В диалоговом окне поместите курсор в поле *Массив данных*. С помощью мыши выделите диапазон значений для функции *Частота*, т.е. ячейки C2:C51. Поместите курсор в поле *Двоичный массив* и с помощью мыши выделите диапазон интервалов, т.е. ячейки D2:D32. Щелкните по кнопке *Готово*. Поместите курсор в строку формул и нажмите Shift + Ctrl + Enter, указав тем самым, что вводится не просто очередная формула, а формула массивов. Результат появится в ячейках E2:E32. Здесь в каждой ячейке помещается количество попаданий результатов в каждую группу.
9. Определите частоту f_i , как n_i/n . Значения частоты для различных интервалов разместите в ячейках F2:F32. Для этого в ячейку F2 запишите формулу $=E2/50$. Скопируйте содержимое ячейки F2 в ячейки F3:F32.
10. Определите среднее арифметическое посева на душу (по формуле 1). В ячейку G2 запишите заголовок «Среднее \bar{x} ». Выделите ячейку H2. Щелкните по кнопке *Мастер функций*. В графе *Категория* выберите *Статистические*. В графе *Функция* выберите *СРЗНАЧ*. В диалоговом окне поместите курсор в поле *Число 1* и с помощью мыши выделите диапазон значений C2:C51.
11. Определите значение дисперсии (по формуле 2). В ячейку G3 запишите заголовок «Дисперсия – $D(x)$ ». Выделите ячейку H3. С помощью *Мастера функций* выберите

функцию *ДИСП*. В качестве диапазона значений укажите C2:C51.

12. Определите стандартное отклонение (по формуле 4). В ячейку G4 запишите заголовок «Стандартное отклонение – $\delta(x)$ ». Выделите ячейку H4. С помощью *Мастера функций* выберите категорию *Математические*, функцию – *КОРЕНЬ*. В качестве *Числа 1* укажите H3.

13. Постройте гистограмму данного закона распределения. Выделите ячейки F2:F32. Вызовите *Мастер диаграмм*. Выберите *Гистограмма / гистограмма 1*. Дайте название диаграммы «Гистограмма», название осей *x* и *f*.

В полученную гистограмму можно вносить изменения. Для этого выделите ее двойным щелчком, затем щелкните правой кнопкой мыши по той области гистограммы, в которую нужно внести изменения. Сначала измените значения оси *x*. Щелкните правой кнопкой мыши в области диаграммы. В появившемся контекстном меню выберите команду *Исходные данные*. В появившемся диалоговом окне - *Исходные данные* щелкните по вкладке *Ряд*. Затем щелкните по окошку *Подписи по оси x*. В качестве диапазона значений для оси *x* задайте ячейки D2:D32. Щелкните по кнопке *Ок*. Для оси *x* будут введены новые значения от 0 до 3 (рис. 3).

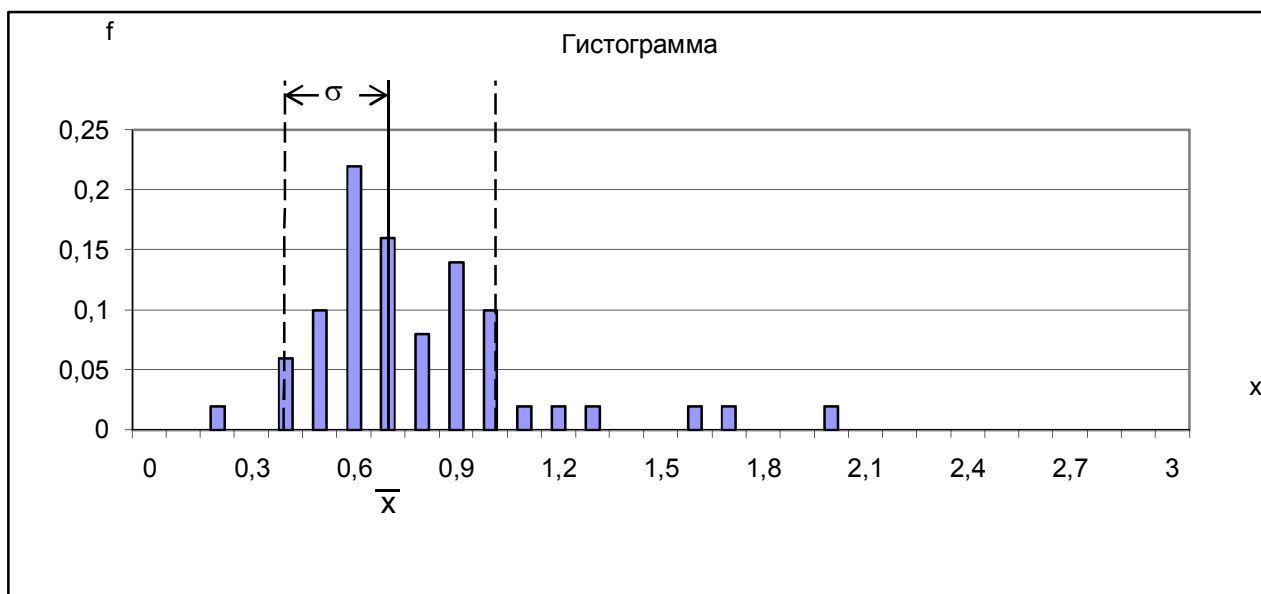



Рис. 3. Гистограмма закона распределения

В полученной гистограмме можно произвести также другие изменения. Два самых крупных объекта гистограммы – это фон, или область диаграммы и пространство между двумя осями, или область построения. Можно изменять цвета, узоры и рамки для этих областей.

Позиционируя курсор между осями, двойным щелчком вы инициируете форматирование области построения; то же самое действие, но вне осей обеспечивает форматирование области гистограммы. Двойным щелчком по столбцам гистограммы можно изменить их раскраску. Отметьте на гистограмме значение среднего \bar{x} и стандартного отклонения δ . Для этого, щелкнув по кнопке *Рисование* , вызовите панель инструментов Рисование. Щелкните по кнопке *Линия* . Курсор превратится в крестик. Этот крестик установите в точку на оси x , соответствующую \bar{x} , нажмите Shift и левую кнопку мыши и перетащите курсор, не отпуская клавиши, до верхней рамки гистограммы. После двойного щелчка на проведенной линии можно изменить ее вид, цвет и толщину.

Точно таким же способом проведите линию через значение $x = \bar{x} + \delta$. Затем при нажатой клавише Ctrl перетащите данную линию к значению $x = \bar{x} - \delta$.

Для размещения на гистограмме дополнительных надписей \bar{x} и δ) необходимо на панели инструментов *Рисование* щелкнуть по кнопке *Надпись* 

Рассматривая построенную гистограмму, можно сделать вывод, что среднее является характеристикой расположения признака x , т.е. посева на душу, представляющий собой типичное значение признака, центральную точку его распределения. По обе стороны от этого значения группируются остальные значения признака.