

Элементы математической статистики



Тема:

*Предмет и задачи математической
статистики. Представление данных.*



Определение

Современную математическую статистику определяют как науку о принятии решений в условиях неопределённости.

Неопределенность не означает незнания, просто она предполагает, что точное определение исхода эксперимента невозможно



Задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Математическая статистика возникла в XVII веке и развивалась параллельно с теорией вероятностей.

Дальнейшее развитие (вторая половина XIX века – начало XX века) обязано, в первую очередь,

П. Л. Чебышеву, А. А. Маркову, А. М. Ляпунову, а так же К. Гауссу, А. Кетле, Ф. Гальтону, К. Пирсону и другие. XX век – советские учёные :

В. И. Романовский, Е. Е. Слуцкий, А. Н. Колмогоров.

Английские: Стьюдент, Фишер, Смирнов.

Американские: С. Нейман, Вальд.



Статистическая совокупность

- Полученные в ходе выполнения ряда независимых опытов значения случайной величины представляют собой простую статистическую совокупность, или статистический ряд, подлежащий обработке и научному анализу.
- *Статистическая совокупность* представляет собой множество объектов, однородных относительно некоторого качественного или количественного признака, характеризующего эти объекты.



В общем смысле под словом **«признак»** подразумевают свойство, проявлением которого один предмет отличается от другого (в биологии: характерные особенности в строении и функциях живого).

- ***Качественные:*** окраска шерстяного покрова, пол животного, тип телосложения, масти лошадей (серые, вороные, гнедые, пёстрые и другие), цвет глаз и волос.

- 
- *Альтернативные признаки:* женщина и мужчина, высокий и низкий.
 - *Количественные признаки* поддаются непосредственному измерению или счёту. Их делят на мерные и счётные.
 - *Мерные признаки:* мясная и молочная продуктивность животных, содержание жира и белка в молоке, живая масса.
 - *Счётные признаки:* плодовитость, яйценоскость и другие.



Генеральная и выборочная совокупности

- Совокупность, состоящая из всех объектов, которые могут быть к ней отнесены на основании однородности некоторого признака, называется **генеральной**.
- Множество объектов, случайно отобранных из генеральной совокупности, называется выборочной совокупностью или **выборкой**. Число объектов выборки называют её объемом и обозначают **n**



Репрезентативность выборки

Для того чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли.

Для того чтобы выборка достаточно хорошо отражала свойства генеральной совокупности, она должна быть представительной (**репрезентативной**). Репрезентативность выборки достигается случайным отбором объектов из генеральной совокупности.

Каждый объект выборки считается отобранным из генеральной совокупности случайно, если все объекты имеют одинаковую вероятность попасть в выборку.

- 
- Существует два основных способа отбора объектов из генеральной совокупности: повторный и бесповторный. Если выборку отбирают по одному объекту, который обследуют и снова возвращают в генеральную совокупность, то выборка является повторной. Если выборка не возвращается в генеральную совокупность, то она является бесповторной.
 - Если в выборку входит до 30 членов ($n < 30$), она называется малой, если более 30 ($n > 30$) – большой.

латинского *variants, variantis* - различимый, изменяющийся)

Характерным свойством биологических признаков является **варьирование** величины признаков в определённых пределах при переходе от одной единицы наблюдений к другой. Эти колебания величины одного и того же признака, наблюдаемые в массе однородных членов статистической совокупности, называют **вариациями** (от латинского *variatio* — изменение, колебания), а отдельные числовые значения варьирующего признака принято называть **вариантами** (от латинского *variants, variantis* - различимый, изменяющийся)

- 
- Для случая, когда количественный признак является дискретной величиной, его значения x_i и соответствующие им частоты n_i или относительные частоты w_i представляют в виде таблицы, в которой значения признака (варианты) располагаются в порядке возрастания.
 - $w_i = \frac{n_i}{n}$, где n_i - это число повторений варианта
 - Такие таблицы называют **статистическим дискретным рядом распределения** или **дискретным вариационным рядом**.



Вариационным рядом или рядом распределения называют двойной ряд чисел, показывающий, каким образом числовые значения признака связаны с их повторяемостью в данной статистической совокупности.

Общая сумма частот вариационного ряда равна объёму данной совокупности, т. е.

$$\sum_{i=1}^k n_i = n$$

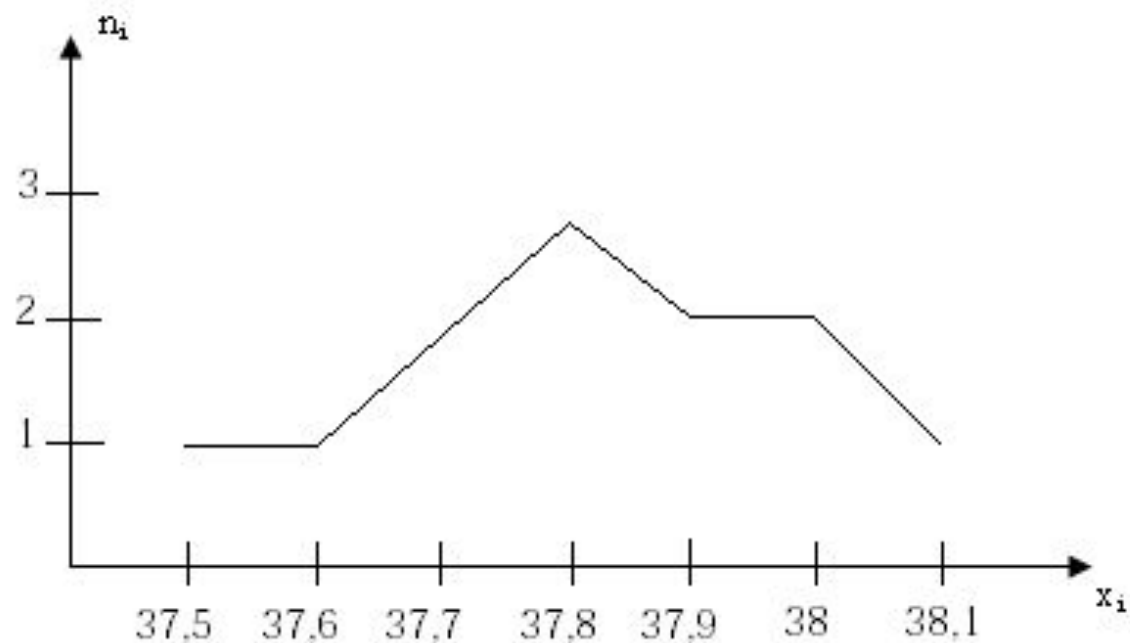


Графически статистические ряды могут быть представлены в виде **полигона, гистограммы или графика накопленных частот (кумулята)**.

- **Полигоном** частот называют ломаную линию, отрезки которой соединяют точки (x_i, n_i) .
- Полигоны обычно служат для изображения выборки в случае дискретной случайной величины.

Пример. В результате измерения температуры у 12 животных получены следующие значения: 37,8; 37,8; 38; 37,7; 37,9; 37,8; 37,5; 37,7; 37,6; 38; 38,1; 37,9. Требуется построить вариационный ряд и соответствующий ему полигон.

X_i	37,5	37,6	37,7	37,8	37,9	38	38,1
n_i	1	1	2	3	2	2	1



- 
- В случае большого количества вариантов и непрерывного распределения признака статистическое распределение признака можно задать в виде последовательности интервалов и соответствующих им частот или относительных частот. Для этого интервал, в котором заключены все наблюдаемые значения признака, разбивают на определенное количество частных интервалов $(x_0; \bar{x}_1]; (x_1; x_2]; \dots; (x_{k-1}; x_k]$ длиной h и находят для каждого частичного интервала сумму частот вариант .
 - Для графического изображения интервального ряда распределения используют **гистограммы**.



Гистограмма относительных частот (или просто гистограмма) - ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h , а высоты равны ω_i / h (или ω_i).

Площадь гистограммы равна единице.

- $f_i = \frac{\omega_i}{h} = \frac{n_i}{n \cdot h}$ - плотность относительных частот.
- $h = \frac{x_{\max} - x_{\min}}{k}$ - длина соответствующего интервала
- $K = 1 + 3,32 \lg(n)$ - количество классов (интервалов)

№	Интервал	Середина интервала		
1	5,05-5,15	5,1	0,05	0,05
2	5,15-5,25	5,2	0,08	0,13
3	5,25-5,35	5,3	0,12	0,25
4	5,35-5,45	5,4	0,2	0,45
5	5,45-5,55	5,5	0,26	0,71
6	5,55-5,65	5,6	0,15	0,86
7	5,65-5,75	5,7	0,1	0,96
8	5,75-5,85	5,8	0,04	1

Гистограмма

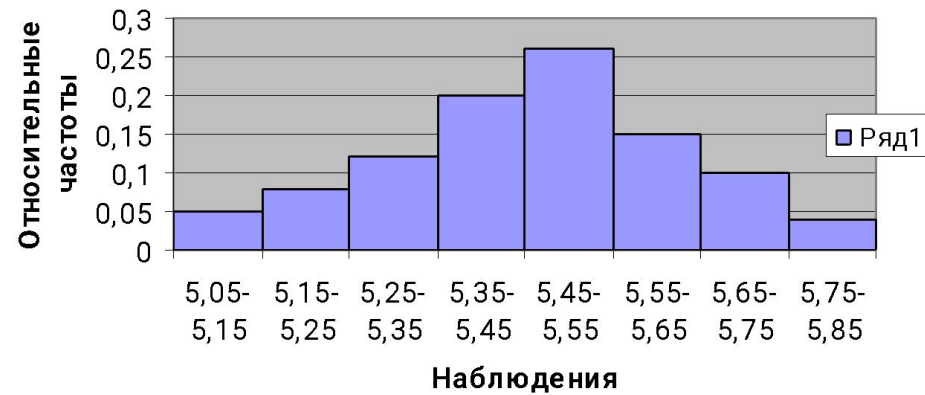
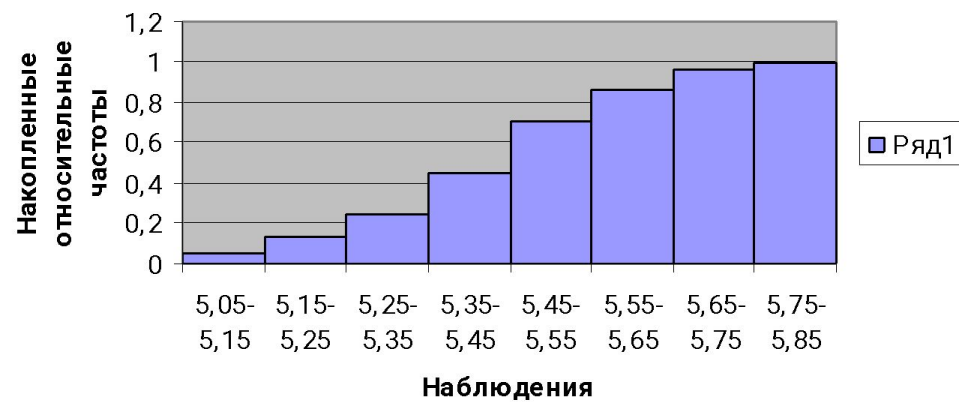


График накопленных частот



- 
- Гистограмму можно рассматривать как график эмпирической (выборочной) плотности распределения $f_i(x)$. Если у теоретического распределения F существует конечная плотность, то эмпирическая плотность является некоторым приближением для теоретической. В этом и состоит практическая польза гистограммы.
 - **Графиком накопленных частот** называется фигура, строящаяся аналогично гистограмме с той разницей, что для расчета высот прямоугольников берутся не простые, а накопленные относительные частоты, т.е.
$$\frac{n_i^h}{n} = \omega_i^h.$$
 - Эти величины не убывают, и таким образом график накопленных частот имеет вид ступенчатой “лестницы” (от 0 до 1).

- 
- **График эмпирической функции распределения проходит через правые верхние углы прямоугольников.**
 - **График накопленных частот и эмпирическая функция распределения на практике используется для приближения теоретической функции распределения.**