**Лекция 1**

1. **Основные теоретические положения**

**1.1 Вариационный ряд, его представление**

**Генеральной совокупностью** называется совокупность всех однородных объектов, подлежащих изучению. Объем генеральной совокупности ***N****.*

**Выборочной совокупностью**, или **выборкой** называется совокупность объектов, случайно отобранных из генеральной совокупности.

**Объемом** выборки ***n*** называется число ее объектов.

Выборка бывает **повторной** (с возвращением исследуемого объекта в генеральную совокупность) и **бесповторной**, при которой отобранный объект в генеральную совокупность не возвращается.

Выборка должна быть **репрезентативной** (представительной), т. е. она должна правильно отображать все свойства и характеристики объектов генеральной совокупности.

Пусть в результате независимых испытаний, проведенных в одинаковых условиях, из генеральной совокупности данных, выбраны числовые значения *x*1, *x*2, … *xn*, характеризующие некоторый признак (*n* ‒ объем выборки). Последовательность наблюдаемых значений *xi* (*i* = 1, 2, …, *n*), записанных в возрастающем порядке, называется **дискретным** **вариационным рядом**, а элементы этой последовательности *xi* называют **вариантами**. Если среди вариант есть равные значения, тогда дискретный вариационный ряд записывается в виде табл. 1, где *n*1, *n*2,…, *nk*– частоты (количества повторений) значений *x*1, *x*2, …, *xk*; при этом должно выполняться условие *n*1 + *n*2 + … + *nk* = *n*.

*Таблица 1*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Значения *Х* | *x*1 | *x*2 | … | *xk* |
| Частоты | *n*1 | *n*2 | … | *nk* |

**Статистическим распределением частот** (распределением частот) называ­ется перечень вариант и соответствующих им частот, записанных в виде таб­лицы.

Относительной частотой  варианты *xi* называется отношение ее частоты к объему выборки: =.

Сумма относительных частот всех вариант равна единице:

.

**Распределением относительных частот** называется перечень вариант и соответствующих им относительных частот.

**Полигоном** **частот** называют ломаную (рис. 1), составленную из отрезков прямых, которые соединяют точки (*x*1, *n*1), (*x*2, *n*2),…, (*xk*, *nk*); здесь *xi* – варианты, *ni* – частоты.

**Полигоном относительных частот** называют ломаную, составленную из отрезков, которые соединяют точки с координатами (,).

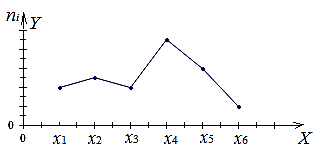


Рис. 1. Полигон частот

**Интервальный вариационный ряд** – это ряд, в котором значения признака могут меняться непрерывно или число значений признака велико. Для построения такого ряда промежуток изменения признака разбивается на ряд отдельных интервалов и подсчитывается количество значений величины в каждом из них. Если все интервалы имеют одинаковую длину, то число интервалов в случае нормально распределенной совокупности рекомендуется находить по **формуле Стерджесса**

*k* = 1 + 3,322 lg(*n*);

обычно, 6 ≤ *k* ≤ 12.

Длина частичного интервала определяется по формуле

*h* = (*xmax* – *xmin*)/*k*;

величина (*xmax*– *xmin*), называется **размахом вариации**. Это разность между наибольшим и наименьшим значениями признака.

**Гистограммой частот** называется ступенчатая фигура (рис. 2), состоящая из прямоугольников с основаниями  и вы-сотами Если интервалы имеют одинаковую длину, то прямо-угольники имеют одно и тоже основание *h*; их высоты равны 

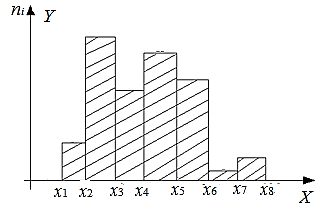


Рис. 2. Гистограмма частот

**Эмпирической функцией распределения**, или **функцией распределения выборки**, называется функция, определяющая для каждого значения *X* относительную частоту события *X* < *x*. Обозначим функцию распределения через *F*\*(*x*); если *nx* – число вариант, меньших *x*, *n* – объем выборки, то по определению

*F*\*(*x*) = *nx* / *n*.

Свойства функции *F*\*(*x*):

1) значения функции *F*\*(*x*) принадлежат отрезку [0; 1];

2) *F*\*(*x*) – неубывающая функция;

3) если *xmin* – наименьшая, *xmax* – наибольшая варианта, то

*F*\*(*x*) = 0, при *x* < *xmin*; *F*\*(*x*) = 1, при *x* *xmax*.

Выборочную функцию распределения можно задать аналитически, таблично (табл. 2) или графически (рис. 3, 4).



*Таблица 2*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Варианты |  |  |  |  |  |
|  |  |  |  |  | 1 |

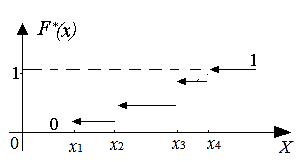


Рис. 3. График функции распределения дискретной функции

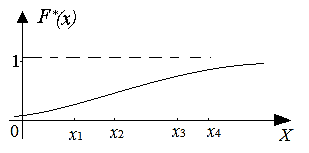


Рис. 4. График функции распределения непрерывной функции

Оценки параметров генеральной совокупности, полученные на основании выборки, называются **статистическими**. Оценку параметра *θ* обозначим через  статистическая оценка является случай-ной величиной и зависит от *x*1, *x*2, … , *xn*, т. е.

 = (*x*1, *x*2, … , *x*n).

* 1. **Числовые характеристики оценок параметров распределения**

Оценка параметра  называется **несмещенной**, если математическое ожидание  равно *θ*, т.е.

M() = *θ*

и **смещенной**, если

M() ≠ *θ*.

Оценка  параметра *θ* называется **состоятельной**, если при любом

ε > 0, .

Оценка  называется **эффективной**, если при заданном объеме выборки *n* она имеет наименьшую дисперсию, т.е. *D*() = *Dmin*.

Оценка  имеет практическую ценность, если она является несмещенной, состоятельной и эффективной.

Если статистическая оценка характеризуется одним числом, она называется **точечно**й. К числу таких оценок относятся выборочная средняя и выборочная дисперсия.

**Выборочной средней** *x*в называется среднее арифметическое значений признака выборочной совокупности. Если все значения *x*1, *x*2, … , *xn* различны, то

.

Если значения *x*1, *x*2, … , *xk*, имеют соответственно частоты *n*1, *n*2, …,*nk*, (*n*1 + *n*2 + …+ *nk* = *n*), то

.

Выборочную среднюю иногда обозначают так: .

Величину принимают в качестве оценки генеральной средней. Можно показать, что эта оценка является несмещенной и состоятельной.

**Выборочной дисперсией**  называют среднее арифметическое квадратов отклонения наблюдаемых значений выборки от их среднего значения *x*в.

Если все значения *x*1, *x*2, … , *xn* признака выборки объема *n* различны, то



Если среди значений выборки есть равные числа, но сами значения {*x*1, *x*2, … , *xk*} различные числа, повторяющиеся с частотами {*n*1, *n*2,…, *nk*}, причем *n*1 + *n*2 + …+ *nk* = *n*, то

.

Выборочное **среднеквадратическое отклонение**  определяется формулой

.

Для вычисления выборочной дисперсии можно также пользоваться формулой

,

где

; .

Выборочная дисперсия оценивает дисперсию генеральной совокупности и является смещенной оценкой. Чтобы получить несмещенную оценку генеральной дисперсии выборочную дисперсию умножают на величину  и получают так называемую **эмпирическую** (или «**исправленную**») дисперсию

;

.

При *n* 50 практически нет разницы между оценками  и . Эти оценки являются состоятельными оценками генеральной дисперсии.

Для оценки среднеквадратического отклонения генеральной совокупности служит **«исправленное» среднеквадратическое отклонение**

.

**Интервальной оценкой** называют числовой интервал, определяемый по результатам выборки, в котором с вероятностью γ находится неизвестное значение параметра генеральной совокупности θ, т.е.

.

Вероятность γ называют **доверительной вероятностью** или **надежностью**; обычно выбирают значения для вероятности γ близкими к единице (0,95; 0,98; 0,99 и т.д.). Интервал  называется **доверительным интервалом**.

Доверительный интервал можно найти из условия



где > 0 – некоторое число. Отсюда следует, что



Эта формула означает, что доверительным интервалом является промежуток .

Доверительный интервал **оценки математического ожидания *a*** **нормального распределения** **при известном значении среднеквадратического отклонения σ**.

; .

При точности оценки  и  получим



Число *t* является аргументом функции Лапласа (Приложение 2).

Доверительный интервал для **оценки математического ожидания *a* нормального распределение при неизвестном значении среднеквадратического отклонения *S*.**

При точности оценки  и значении  получим



Число  является аргументом распределения **Стьюдента (**Приложение 3) с  степенями свободы и уровнем значимости . Распределение Стьюдента определяется параметром *n* и не зависит от неизвестных параметров *a* и .

Доверительный интервал для **оценки генеральной дисперсии** совокупности *X*, распределенной по нормальному закону, покрывающего среднеквадратическое отклонение  с надежностью .

Введем параметр . Тогда получим условие



Рассмотрим случайную величину , которая имеет **распределение Пирсона,** зависит только от *n* и не зависит от оцениваемого параметра . Вероятность для параметра  равна



Из этого уравнения по табл. Приложения 5 для распределения Пирсона (– распределения) можно найти параметр *q* или используем табл. Приложения 4. Получив значение *q*, найдем доверительный интервал

.

Если  то доверительный интервал вычисляют так:

