**Лекция 1**

**Вероятностные основы статистического анализа**

*Вероятностное пространство, аксиоматика А.Н. Колмогорова, случайная величина и ее числовые характеристики, основные законы распределения случайной величины, закон больших чисел.*

Вероятностное пространство определяется тройкой . Здесь -пространство элементарных событий, это множество элементарных событий таких, что выполняются два условия:

1) в результате реализации эксперимента всегда происходит одно из элементарных событий ;

2) все взаимно исключают друг друга.

Пространство элементарных событий считается заданным, если указаны все его элементы. Это пространство может быть конечным или счетным.

Любое подмножество пространства элементарных событий, то есть , является событием, если оно входит в класс подмножеств множества , замкнутых относительно основных теоретико-множественных операций – объединение, пересечение и дополнение. При этом выполняются следующие три условия:

1) ,

2)

3)

Из первых двух условий следует, что . Из условий 1)-3) следует, что класс

представляет собой -алгебру.

Вероятностная мера определяется как числовая функция . При этом справедливы аксиомы А.Н. Колмогорова:

1. .

Случайная величина – измеримая функция, отображающая пространство элементарных событий во множество действительных чисел, для которой прообраз любого борелевского множества есть -алгебра.

Законы распределения случайной величины определяются в соответствии со способами определения функции: для дискретной случайной величины в виде таблице, где каждому возможному значению случайной величины соответствует своя вероятность, а для непрерывной случайной величины вводится понятие функции распределения как вероятность того, что случайная величина принимает значения, меньше некоторого заданного.

Закон больших чисел описывается в виде теорем – неравенство Чебышева, теорема Чебышева и теорема Бернулли. Эти теоремы характеризуют сходимость по вероятности среднего арифметического случайных величин к среднему арифметическому их математических ожиданий. В случае теоремы Бернулли это выражается как сходимость по вероятности относительной частоты наступления события к вероятности наступления этого события при одном испытании Бернулли.

Закон больших чисел в форме Чебышева описывается следующим образом. Неравенство Чебышева.

Если  на  и ,то 

Пусть ,то .

Неравенство Чебышева. Для любой случайной величины  имеющей конечную дисперсию ,при каждом  имеет место неравенство.

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ВЫБОРКИ

Пусть случайный эксперимент описывается случайной величиной *Х.* Повторяя случайный эксперимент *n* раз, получим последовательность наблюденных значений , ,…, случайной величины *Х*, называемых выборкой из генеральной совокупности , описываемой функцией распределения *F (x).*

Рассмотрим основные характеристики эмпирического распределения случайной величины *Х.*

*Средним арифметическим наблюденных значений выборки называется величина, определяемая по формуле*

, (6.1)

где - наблюденное значение с частотой ; *n* - число наблюдений, . Частоты могут быть равны единице, , тогда *k=n.*

Согласно закону больших чисел, при неограниченном увеличении числа наблюдений среднее арифметическое сходится по вероятности к математи-ческому ожиданию случайной величины *X*.

В качестве числовой характеристики выборки , ,…, применяется *медиана* . Чтобы вычислить ее, все наблюдения располагают в порядке возрастания или убывания. При этом, если число вариант нечетно, т.е. 2*m+*1 то медианой является *m+*1варианта = ); еслти же число вариант четное, то медиана равна среднему арифметическому двух средних значений:

.

Наблюдение выборки, имеющее наибольшую частоту, называется модой и обозначается . Среднее арифметическое , медиана , мода — это средние значения выборки.

Для описания рассеивания наблюденных значений случайной величины Х относительно среднего арифметического используются статистическая дисперсия и среднее квадратическое отклонение.

*Статистической дисперсией выборочного распределения называется среднее арифметическое квадратов отклонений значений наблюдений от средней арифметической :*

, (6.2)

где - наблюденное значение с частотой ; - число наблюдений.

*Средним квадратическим отклонением или стандартной ошибкой называется квадратный корень из статистической дисперсии:*

. (6.3)

Как следует из теоремы 2.5, статистическую дисперсию можно вычислять по формуле

. (6.4)

Действительно,

.

Средние величины, характеризуя выборочное распределение одним числом, не учитывают вариацию случайной величины *X*, связанной со случайным экспериментом. В математической статистике используется ряд способов для определения вариации.

*Вариационный размах R (или широта распределения) есть разность между наибольшим и наименьшим значениями выборочного распределения: R = .*

Вариационный размах используется при статистическом изучении качества продукции.

Средним линейным отклонением называется среднее арифметическое модулей отклонений значений выборки от средней:

(6.5)

*Коэффициентом вариации по среднему квадратическому (среднему линейному) отклонению называется отношение среднего квадрати-ческого (среднего линейного) отклонения к средней, выраженное в процентах (или в долях единицы):*

. (6.6)

Коэффициент вариации служит для сравнения двух эмпирических распределений с точки зрения их рассеивания относительно среднего арифметического.

Обобщающими характеристиками выборочных распределений являются статистические моменты распределения.

1. Начальные статистические моменты *k-*го порядка

(6.7)

Тогда:

при k = 0

при k = 1

при k = 2

при k = 3

при k = 4 и т.д.

Практичеси используются моменты первых четырех порядков.

2. Центральные статисчитеские моменты порядка

Тогда:

при k = 0 ;

при k = 1 ;

при k = 2

при k = 3

при k = 4 и т.д.

Отметим, что центральный статисчитеский момент третьего порядка служит мерой асимметрии распределения выборки. Если распределение симметрично, то .

На практике моменты порядка выше четвертого почти не применяются, так как обладают очень высокой дисперсией и из сколько-нибудь надежное определение потребовало бы выборок большого объема.

С помощью бинома Ньютона легко установить следующие формуля для вычисления третьего и четвертого центральных статисчитеских моментов через начальные статистические моменты:

;

Действительно, например,

Характеристика асимметрии выборочного распределения вычисляется по формуле , а эксцесс выборочного распределения определяется характеристикой .