**Лекция 7.**

**Выборочные статистики:** *Математическая модель эксперимента, генеральная совокупность, выборка, относительная частота, статистический ряда (дискретный и интервальный), формула Стэрджеса, распределения – дискретный и интервальный, полигон и гистограмма, эмпирическая функция распределения, основные характеристики выборки*

Рассмотрим множество или, как мы будем говорить, *генеральную* *совокупность* из *n* элементов *а1, а2, ..., аn.* Произвольное упорядоченное множество  из rэлементов, входящих в генеральную совокупность, будем называть *выборкой объема r* из этой генеральной совокупности. Наглядно мы можем просто представить себе, что элементы выбираются один за другим. Возможны два способа выбора. Первый – *выбор с возвращением;* в этом случае производится каждый раз из всей генеральной совокупности, и тот же элемент может быть выбран более одного раза. В выборке могут встречаться повторения. Второй – *выбор без возвращения.* При таком способе выбора однажды выбранный элемент удаляется из генеральной совокупности, так что выборка не содержит повторяющихся элементов. Очевидно, что в этом случае объем выборки не может превысить объема генеральной совокупности *n.*

При выборе с возвращением каждый из *r* членов выборки быть выбран nразличными способами. Поэтому, применив предыдущую теорему при n1=n2=…=n*,* получаем, что число различных возможных выборок равно *пr.* В случае выборок без возвращения первый член выборки может быть выбран nспособ ми, второй n-1 третий  *n-1* и т. д. Воспользовавшись той жетеоремой, получим что общее число выборок равно n(n– 1)... (n-r+1) .

Если nвелико, а *r* сравнительно мало, то отношение *(п)r/пr* близко к единице. Это говорит о том, что при большой генеральной совокупности и относительно малой выборке различие между двумя способами выбора незначительно.

Мы ввели терминологию, взятую из практики, но не говорили еще о применимости модели случайного выбора к действительным опытам. Бросание монеты или кости и другие аналогичные действия могут быть истолкованы как опыты, практически дающие случайный выбор с возвращением, и наши вероятности численно близки к частотам, наблюдаемым в длинной серии опытов, несмотря даже на то, что абсолютно симметричной монеты или кости не существует. Типичным примером случайного выбора без возвращения может служить вытаскивание карт из перетасованной колоды (причем предполагается, что колода тасуется гораздо тщательнее, чем это делается обычно). При выборе из совокупностей, состоящих из людей, статистики встречаются со значительными и часто неожиданными трудностями, и опыт показывает, что очень трудно получить даже грубое приближение к случайной выборке.

Для статистической оценки рассматривается случайная выборка объёма r. Из генеральной совокупности *n* элементов *а1,а2,…,аn* извлекается с возвращением выборка объёма r. Рассмотрим событие А, состоящее в том, что все элементы выборки различны, иными словами, что наша выборка могла бы быть получена при выборе без возвращения. Последняя теорема показывает, что различных выборок с возвращением существует *nr*, из них *(n)r* обладают указанным свойством. Приписывая всем точкам пространства элементарных событий равные вероятности, получаем, что вероятность отсутствия одинаковых элементов в выборке (вероятность события А) равна

 (3.1)

Интересные свойства этой формулы вскрываются следующими конкретными её интерпретациями.

Случайные числа возникают при анализе равномерного распределения. Допустим, что генеральная совокупность состоит из десяти цифр:0,1,2,…,9. Любая последовательность из пяти цифр представляет выборку объёма *r=*5. Свяжем с каждой такой последовательностью вероятность 10-5. Согласно (3.1), вероятность того, что пять последовательных случайных цифр различны, равна *p*=(10)510-5=0.3024.

Интуиция подсказывает нам, что в математических таблицах, содержащих величины, вычисленные с большим числом десятичных знаков, последние пять знаков должны обладать многими свойствами случайных чисел.

(В обычных логарифмических и др. таблицах с малым числом знаков разность между случайными величинами близка к постоянной (из-за непрерывности), и поэтому последние знаки изменяются закономерно). Для опыта выбраны 16-значные таблицы и было сосчитано, для скольких чисел пять последних знаков различны. В первых 12 сериях по 100 чисел частоты наступления этого события оказались следующими:

0,30 0,27 0,30 0,34 0,26 0,32

0,37 0,36 0,26 0,31 0,36 0,32

взяли первые 100 чисел в таблице и сосчитали, сколько в них последних пять знаков различны -= оказалось 30. Далее идёт следующая серия . . . и так до 12.

Эти частоты колеблются около 0,3024, и методы математической статистики показывают, что амплитуда этих колебаний не превышает ожидаемой. Средняя частота оказывается равной 0,3142, т.е. довольно близкой к вероятности 0,3024.

Рассмотрим теперь число *е*=2,71828 ... . Из первых 800 десятичных знаков этого числа образуем 160 групп по пяти знаков в каждой; эти группы объединим в 16 серий по 10 групп в каждой серии. В этих сериях числа групп, в которых все пять знаков различны, окажутся следующими:

3, 1, 3, 4, 4, 1, 4, 4,

4, 2, 3, 1, 5, 4, 6, 3,

Как видно, частоты колеблются вокруг 0,3024, и теория подтверждает, что размах этих колебаний не больше, чем нужно было ожидать. Средняя частота появления группы в пять различных цифр в 160 группах равна 52/160=0,325, т.е. близка к *р*=-0,3024.

*Если n шаров размещаются по n ящикам, то вероятность того, что каждый ящик будет занят, равна .* Эта вероятность неожиданно мала: при n=7 она равна лишь 0,00612… . Отсюда следует, например, что если в городе каждую неделю происходит семь автомобильных катастроф, то (в предположении, что все возможные распределения равновероятны) практически каждая неделя будет содержать дни, когда происходят две и более катастрофы. Равномерное распределения числа катастроф встречается в среднем лишь для одной недели из 165. Этот пример иллюстрирует неожиданное свойство чистой случайности. При n=6 вероятность n!n-n равна 0,01543… . это показывает, как мала вероятность того, что при шести бросании правильной кости появятся все грани.