

## Раздел 2

### ВВЕДЕНИЕ В МАТЕМАТИЧЕСКУЮ СТАТИСТИКУ

В настоящее время математическая статистика широко используется практически в каждой отрасли человеческой деятельности.

Методы математической статистики используются при принятии решения в условиях неопределенности. Для ситуации неопределенности характерно то, что по состоянию в настоящий момент невозможно с точностью предсказать состояние в будущем некоторого процесса или явления, несмотря на то, что факторы, влияющие на данный процесс, не меняют своего значения.

Математическая статистика работает с данными, которые получаются из проводимых наблюдений. Эти наблюдения проводятся с целью получения заключения об источнике наблюдения, который характеризуется всеми мыслимыми объектами.

**Пример.** Спонсор телевизионной программы хочет выяснить популярность своей программы по сравнению с другими.

Для решения этого вопроса необходимо провести опрос телезрителей в тот час, когда идет данная программа. Источником наблюдений являются все телезрители в данное время. Если бы опросили всех или почти всех телезрителей и определили процент тех, кто смотрит эту передачу, то вопрос был бы решен.

Но на практике такое не представляется возможным. Опрашивается только какая-то часть телезрителей, и на основании этого необходимо сделать вывод.

Статистические методы позволяют сделать вывод с определенной достоверностью (вероятностью) по тем наблюдениям, которые получены и которых значительно меньше, чем всех, которые могли бы быть сделаны (мыслимых).

Все мыслимые объекты называются *генеральной совокупностью*, количество которых  $N$  может быть достаточно большим ( $N \rightarrow \infty$ ).

Статистические методы используют данные, полученные с некоторого числа объектов, выбранных из генеральной совокупности случайным образом, т. е. таким образом, что каждый объект генеральной совокупности имеет один и тот же шанс быть выбранным.

На этих объектах, выбранных из генеральной совокупности, измеряется или подсчитывается значение показателя, характеризующего некоторое свойство, обозначаемое  $X$ , а результаты, полученные на объек-

тах, обозначаются  $x_1, x_2, \dots, x_n$ , где  $n$  – число объектов.

Наблюдения  $x_1, x_2, \dots, x_n$ , полученные на объектах из некоторой генеральной совокупности, называются *выборкой объема  $n$* .

Методы, касающиеся сбора и обработки данных, называются *методами описательной статистики*.

Часть статистических методов, позволяющих сделать выводы о генеральной совокупности по полученной выборке, называется *статистическими выводами*.

Основная задача статистических выводов – получение заключения об источнике наблюдения по выборке с определенной степенью уверенности, которая оценивается вероятностью.

*Требования к выборке.* Для того чтобы сделать правильный вывод о генеральной совокупности по выборке, выборка должна быть *репрезентативной*, т. е. правильно представлять генеральную совокупность. Выборка будет обладать таким свойством, если каждый объект генеральной совокупности будет иметь один и тот же шанс быть выбранным, в этом случае выборка является *случайной*.

Выборки делятся относительно их размера.

Если выборка содержит меньше 30-ти элементов, то она называется *малой*.

Если выборка содержит больше 30-ти элементов, то она называется *большой*.

Выбор статистических методов зависит от того, какого объема выборка. Для выборок малого объема необходимо выбирать специально разработанные методы.

По уровню наблюдения выборочные данные делятся:

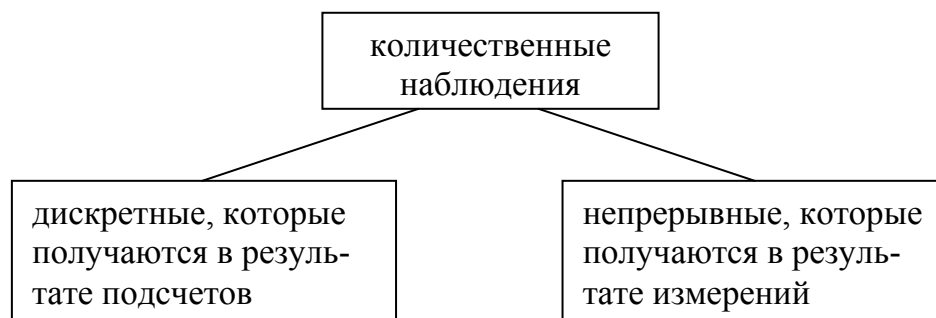
- на качественные;
- количественные.

Качественные данные представляются (кодируются) определенным числом в соответствии с некоторым свойством.

Например, фиксируется пол новорожденных в каком-либо районе города, при этом рождение девочки отмечается «1», рождение мальчика «0» или оценивается качество радиоаппаратуры по шкале: «превосходно», «отлично», «хорошо», «удовлетворительно» (что соответствует, например, следующим баллам: 6, 5, 4, 3).

В дальнейшем будет предполагаться, что наблюдения будут представляться количественной информацией. При этом следует выделить:

- непрерывный тип наблюдений;
- дискретный тип наблюдений.



Случайная переменная *дискретного* типа может принимать лишь изолированные значения на некотором интервале. Случайная переменная *непрерывного* типа – любое значение из некоторого интервала.

К *непрерывным* относятся переменные, значения которых получены в результате измерений (иногда они могут быть округлены и представлены в виде целых чисел). Такие переменные, как вес, рост, температура, расстояние, время и т. д., являются непрерывными.

Для работы с качественным уровнем данных используются специально разработанные для этого уровня данных методы, которые называются *непараметрическими*. Следует отметить, что для количественного уровня данных могут использоваться непараметрические методы. Но методы, разработанные для количественного уровня данных, или *параметрические* методы, не могут использоваться для качественного уровня данных.