Лекция – 2 (2ч)

Тема: Математическая статистика, её основные понятия и приложения к физической культуре и спорту.

План:

1. Возникновение и развитие математической статистики
2. Статистические данные
3. Статистические признаки, совокупности
4. Кривая нормального распределения
5. Соответствие нормальному закону распределения
6. Виды представления статистических данных

1. Возникновение и развитие математической статистики

Издавна в каждом государстве соответствующими органами власти собирались сведения о числе жителей по полу, возрасту, занятости в различных сферах труда, наличии различных воинов, вооружения, денежных средств, орудий труда, средств производства и т.д. Все эти и подобные им данные называются статистическими.

С развитием государства и международных отношений возникла необходимость анализа статистических данных, их прогнозирование, обработка, оценка достоверности основанных на их анализе выводов и т.п. К решению таких задач стали привлекаться математики. Таким образом, в математике сформировалась новая область — математическая статистика, изучающая общие закономерности статистических данных или явлений и взаимосвязи между ними.

Сфера применения математической статистики распространилась во многие, особенно экспериментальные, науки. Так появились экономическая статистика, медицинская статистика, биологическая статистика, статистическая физика и т.д. С появлением быстродействующих ЭВМ возможность применения математической статистики в различных сферах деятельности человека постоянно возрастает. Расширяется ее приложение и к области физической культуры и спорта. В связи с этим основные понятия, положения и некоторые методы математической статистики рассматриваются в курсе “Спортивная метрология”. Остановимся на некоторых основных понятиях математической статистики.

2. Статистические данные

В настоящее время под термином "статистические данные" понимают все собранные сведения, которые в дальнейшем подвергаются статистической обработке. В различной литературе их еще называют: переменные, варианты, величины, даты и т.д. Все статистические данные можно разделить на:

качественные, труднодоступные для измерения (имеется, не имеется; больше, меньше; сильно, слабо; красный, черный; мужской, женский и т.д.)

количественные , которые можно измерить и представить в виде числа общих мер (2кг, 3м, 10раз, 15с и т.д.);

точные , величина или качество которых не вызывают сомнений (в группе 6 человек, 5 столов, деревянный, металлический, мужской, женский и т.д.), и

приближенные, величина или качество которых вызывает сомнение (все измерения: рост 170 см, вес 56 кг, результат бега на 100 м - 10,3 с и т.д.; близкие понятия — синий, голубой, мокрый, влажный и т.д.);

определенные (детерминированные), причины появления, не появления или изменения которых известны (2 + 3 = 5, подброшенный вверх камень обязательно будет иметь вертикальную скорость, равную 0 и т.д.), и

случайные, которые могут появляться и не появляться или не все причины изменения которых известны (пойдет дождь или нет, родится девочка или мальчик, команда выиграет или нет, в беге на 100 м — 12,2 с, принятая нагрузка вредна или нет). В большинстве случаев в физической культуре и спорте мы имеем дело с приближенными случайными данными.

3. Статистические признаки и совокупности

Общее свойство, присущее нескольким статистическим данным, называют их ***статистическим признаком***. Например, рост игроков команды, результат бега на 100 м, принадлежность к виду спорта, частота сердечных сокращений и т.д.

Статистической совокупностью называют несколько статистических данных, объединенных в группу хотя бы одним статистическим признаком. Например, 7.50, 7.30, 7.21, 7.77 — результаты прыжка в длину в метрах у одного спортсмена; 10, 12, 15, 11, 11 — результаты подтягивания на перекладине пяти студентов и т.д.

Число данных в статистической совокупности называют ее ***объемом*** и обозначают **n** .

Различают следующие совокупности:

бесконечные — $n \to \infty$ (масса планет Вселенной, число молекул и т.д.);
конечные — n - конечное число;
большие — $n > 30$;
малые — $n \leq 30$;
генеральные — содержащие все данные, обусловленные постановкой задачи;
выборочные — части генеральных совокупностей.

Например, пусть рост студентов 17−22 лет в РФ — генеральная совокупность, тогда рост студентов КГАФК, всех студентов города Краснодара или студентов II курса — выборки.

4. Кривая нормального распределения

При анализе распределения результатов измерений всегда делают предположение о том распределении, которое имела бы выборка, если бы число измерений было очень большим. Такое распределение (очень большой выборки) называют распределением генеральной совокупности или *теоретическим*, а распределение экспериментального ряда измерений — *эмпирическим*. Теоретическое распределение большинства результатов измерений описывается формулой нормального распределения, которая впервые была найдена английским математиком Муавром в 1733 г.:

Это математическое выражение распределения позволяет получить в виде графика кривую нормального распределения, которая симметрична относительно центра группирования (обычно это значение, моды или медианы). Эта кривая может быть получена из полигона распределения при бесконечно большом числе наблюдений и интервалов. Заштрихованная область графика на рисунке 3 отражает процент результатов измерений, находящихся между значениями x1 и x_2.

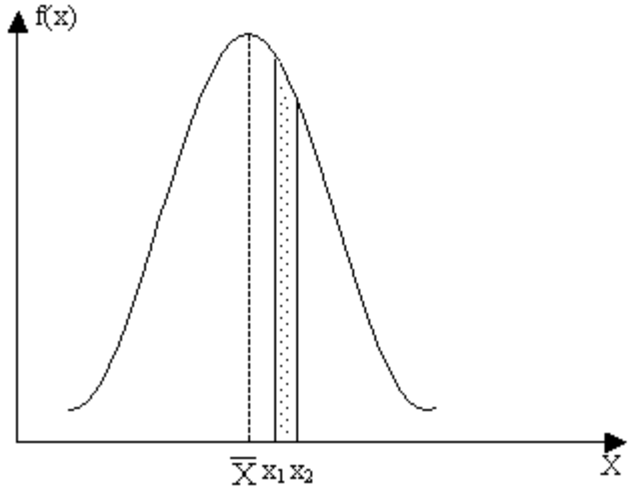


Рис. 3. Кривая нормального распределения.

Введя обозначение $u = \frac{x - \overline{X}}{\sigma}$, которое называется *нормированным* или *стандартизованным* отклонением, получают выражение для нормированного распределения:

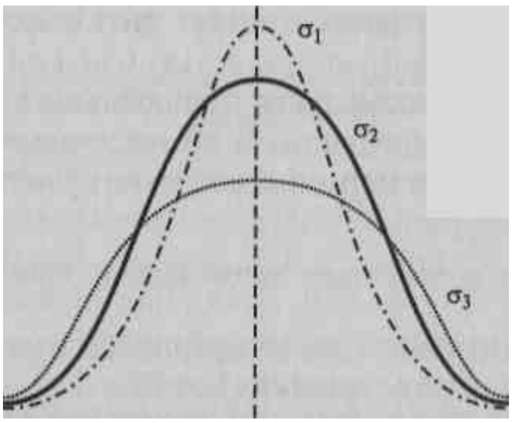
Нормальный закон распределения

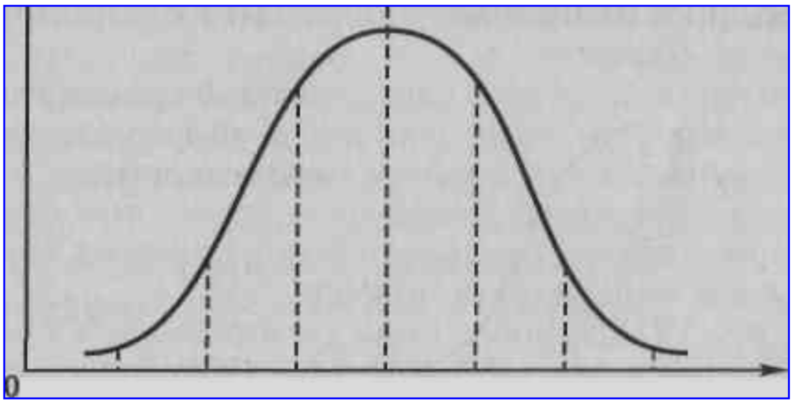
Распределение представляет собой соотношение элементов совокупности с частотой их появления. Соотношение может быть представлено графиком или выражено формулой. В простейшем представлении распределение выглядит как полигон.

Самым распространенным, хорошо изученным и практически полезным распределением принято считать нормальное распределение, отражаемое ***кривой Гаусса*** и известное как ***нормальный закон***. Основой этого распределения является ***закон больших чисел***, доказанных теоремой А.А. Ляпунова.

Закон больших чисел рассматривает ситуацию, при которой утверждается, что с любой вероятностью близкой к 1, отклонение средней арифметической достаточно большого числа случайных величин от некоторой постоянной величины не превзойдет заданного, как угодно малого положительного числа.

Идея нормального распределения в том, что множество единиц совокупности распределяется таким образом, чтобы около средней арифметической было сконцентрировано наибольшее количество единиц, все прочие единицы должны соответствовать кривой Гаусса.





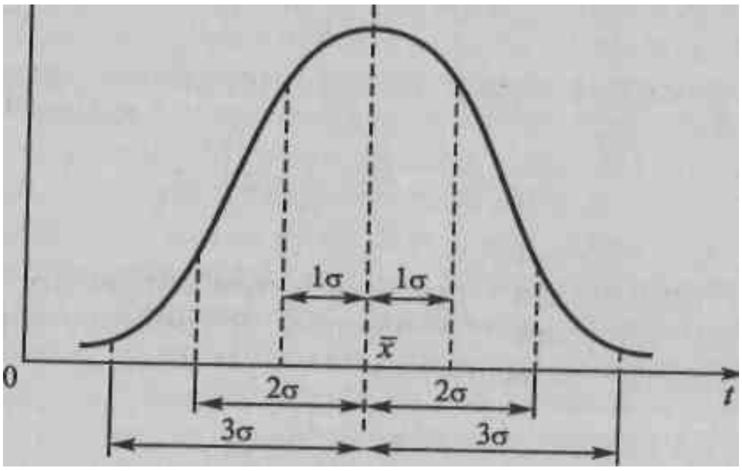
Свойства нормального закона

1. От величины средней арифметической $\overline{X}$ зависит положение кривой: с увеличением (уменьшением) кривая будет сдвигаться вправо (влево) вдоль оси абсцисс, при этом форма её не будет изменяться.
2. От величины среднего квадратического отклонения σ зависит форма кривой: чем больше квадратическое отклонение σ, тем ниже и шире кривая; чем меньше, тем выше и тоньше кривая.

5. Соответствие нормальному закону распределения

С помощью нормального закона решается множество статистических задач. Соответствие нормальному закону на практике решается двумя путями: 1) критериями согласия; 2) правилом трех сигм.

Самым распространенным и простым является критерий χ^2 Пирсона.



Идея критерия состоит в следующем: определяют такие теоретические частоты, которые соответствуют нормальному закону распределения, затем их сравнивают с реальным распределением в соответствии с критерием χ^2 Пирсона. Результаты сравнения показывают, соответствуют ли частоты реального распределения нормальному закону.

Правило трех сигм ($\pm 3\sigma$) является способом проверки эмпирического ряда на соответствие нормальному закону распределения.

Установлено, что под кривой Гаусса участок $\overline{X} \pm \sigma$ занимает 0,6828 всей площади, участку $\overline{X} \pm 2\sigma$ отведено 0,9545 сей площади, а на участке $\overline{X} \pm 3\sigma$ находится 0,9973 всей площадию.

Для оценки варьирования результатов измерений используют следующие соотношения:

± 1,96 σ(u= ± 1,96)	интервал включает	95%	всех результатов
± 2,58 σ(u= ± 2,58)		99%	
± 3,29 σ(u= ± 3,29)		99,9%	
± 1σ (u= ±1)		68,27%	
± 2σ(u= ±2)		95,45%	
± 3σ(u= ±3)		99,73%	

Другими словами, отклонения, большего, чем σ, следует ожидать примерно в одном случае из трех; отклонения, большего, чем 2 σ, — в четырех-пяти случаях из 100 отклонения; большего, чем 3 σ, — в трех из 1000. Последнее соотношение для нормального распределения называют "правилом трех сигм" и используют при исключении сильно отклоняющихся "ошибочных" результатов измерений.

6. Виды представления статистических данных

После того, как определена выборка и стали известны ее статистические данные (варианты, даты, элементы и т.д.), возникает необходимость представить эти данные в удобном виде для решения задачи. На практике используют много различных видов представления статистических данных. Наиболее часто употребляют следующие:

а) табличный вид;

б) вариационный ряд;

в) графический вид.

Если при статистической обработке совокупности безразлично в какой последовательности записывать данные, то бывает удобным расположить эти данные (варианты) в соответствии с их значением либо по возрастанию x_i ~ 2, 3, 3, 5, 5, 6, 6, 6, 6, 7 (неубывающая совокупность), либо по убыванию x_i ~ 7, 6, 6, 6, 6, 5, 5, 3, 3, 3, 2 (невозрастающая совокупность). Этот процесс называется ***ранжированием***. А место каждой варианты в ранжированном ряду называется ***рангом***.