Математическая статистика и её основные понятия.

**Математическая статистика** – часть математики, изучающая особенности и общие закономерности групп из двух и более случайных данных, объединённых общим свойством. **Случайное данное** – объект, который может появиться или не появиться и не все причины его появления или не появления известны. **Статистические данные** - случайные данные, предназначенные для анализа, **Статистический признак** - общее свойство, объединяющее несколько статистических данных в группу – **статистическую совокупность**. **Качественные** статистические данные – выраженные словами без использования единиц измерения, например: сильный, большой, дождливый, красный т. д. **Количественные** статистические данные – представленные числовыми значениями единиц измерения. **Качественная совокупность** – состоит из качественных данных. **Количественная совокупность** – состоит из количественных данных. **Статистическая гипотеза -** предположение, которое проверяется с помощью статистических данных.  **Объём статистической совокупности – п** – количество данных в ней. Примем в дальнейшем - статистические данные обозначать малыми латинскими буквами xi, уi, zi и т. д. с индексом - i, показывающим номер места каждого статистического данного в совокупности и последовательно принимающим все целые числа от 1 до п. Например, имеем совокупность xi кг: 55,73,64,48,67. Здесь п = 5, х3 =64 кг – стоит на третьем месте, начиная от х1 = 55 кг. В количественной совокупности выделяют: **минимальное значение**: xmin; **максимальное значение**: xmax; **размах**: R = xmax – xmin (1). Например, в совокупности xi кг: 55,73,64,48,67 xmin = 48 кг, xmax = 73 кг, R = 25 кг. Любая часть статистической совокупности, составленная по принципу случайного отбора, – её **выборочная совокупность** или **выборка.** Статистические совокупности или их выборки, имеющие 2 ≤ п ≤ 30 – **малые**, если п ˃ 30, то они **большие**. **Ранжированная совокупность** – количественная совокупность, в которой данные по их величине расположены или в неубывающем порядке – каждое следующее данное больше или равно предыдущему ( от xmin до xmax ), или в невозрастающем порядке - каждое следующее данное меньше или равно предыдущему ( от xmax до xmin ). **Ранг количественного данного** есть номер его места в ранжированной совокупности, начиная с первого. При этом, одинаковым по величине данным присваивается одинаковый ранг, равный среднему арифметическому значению из мест, занятых ими в ранжированной совокупности. Пример 1. Ранжированную в неубывающем порядке совокупность xi: 5,7,8,8,8,9,11,12,12 можно заменить совокупностью ri: 1;2;4;4;4;6;7;8,5;8,5, где xi – данные, ri – их ранги, r3= r4= r5 =(3+4+5)/3=4; r8= r9 =(8+9)/2=8,5. Аналогично, ранжированную в невозрастающем порядке совокупность xi: 12,12,11,9,8,8,8,7,5 можно заменить совокупностью r i: 1,5;1,5;3;4;6;6;6;8;9, где r1= r2 = (1+2)/2=1,5; r5= r6= r7 = (5+6+7)/3=6. В одной задаче порядок ранжирования, неубывающий или невозрастающий, для всех совокупностей лучше сохранять одинаковым. Для проверки правильности замены данных их рангами можно воспользоваться формулой Σ r i = п\*(п+1)/2 (2), где Σ r i - общая сумма рангов (заглавную букву Σ-сигма греческого алфавита принято в математике употреблять вместо слова сумма). В рассмотренном выше примере ранги присвоены правильно т.к. Σ r i =1+2+4+4+4+6+7+8,5+8,5=1,5+1,5+3+4+6+6+6+8+9=9\*10/2=45. **Генеральная** статистическая совокупность состоит из всех без исключения статистических данных, соответствующих условию поставленной задачи. По объёму она может быть конечной ( малой или большой) и бесконечной. **Вероятность Р случайной величины х** - Р(х) есть мера возможности её появления или не появления и значения которой могут изменяться от ноля до единицы или, соответственно, от ноля процентов до ста процентов, т.е. 0 ≤ Р(х) ≤ 1 (3) или 0% ≤ Р(х)(3’). При этом: если Р(х) равна нулю или нулю процентов, то появление х невозможно; если Р(х) равна единице или ста процентам, то появление х обязательно; вероятность появления случайной величины Р1 плюс вероятность её не появления Р1 всегда равно единице или ста процентам, т.е. Р1 (х) + Р2 (х) =1 (4) или Р1 (х) + Р2 (х) =100% (4’). Если известно число возможных появлений п случайного события х из общего числа N случайных событий, содержащих события х, то Р(х) = n / N (5) или Р(х) = (n / N)\*100% (5’) – теоретическая вероятность [1]. Например, вероятность вынуть из 36-карточной колоды бубновую карту, которых в колоде 9 штук, будет: Р(х) = 9/36 = ¼ = 0,25 или Р(х) = (9/36)\*100% = 25%. Формулы (4,4’, 5,5’) справедливы и для экспериментального определения вероятности случайного события х, если оно появилось п раз из N равновозможных испытаний для его появления – экспериментальная вероятность [1]. Например: монета подбрасывается вверх 20 раз (N =20) и падает на пол вверх орлом 8 раз (п1=8), вверх решкой 12 раз (п2=12). В этом случае вероятность появления орла Р1 (х)=8/20=0,4 или Р1 (х)=(8/20)\*100%=40%, а вероятность появления решки Р2 (х) =12/20=0,6 или Р2 (х) =(12/20)\*100%=60%. Очевидно, что по (4,4’) 0,4+0,6=1 или 40%+60%=100%. Аналогично: баскетболист 10 раз бросает мяч в кольцо с одного места площадки (N =10) и попадает в кольцо 7 раз (п=7). Отсюда, вероятность попадания мяча в кольцо Р(х) =7/10=0,7 или Р(х) =(7/10)\*100%=70%. Приведённое выше понятие вероятности распространяется и на вероятность справедливости противоположных статистических гипотез. Например, если вероятность уверенности в справедливости одной из них β, то вероятность в справедливости другой будет α = 1 – β или α %=100%- β %. **Уровень значимости α (или р) –** вероятность неуверенности в правильности полученных выводов относительно рассматриваемой статистической гипотезы. В математико-статистических таблицах приняты 3 уровня значимости: 1 уровень: α = 0,05 или α = 5%; 2 уровень: α = 0,01 или α = 1%; 3 уровень: α = 0,001 или α = 0,1%.  **Доверительная вероятность β –** вероятность уверенности в правильности полученных выводов относительно рассматриваемой статистической гипотезы. Аналогично уровню значимости в математико-статистических таблицах приняты 3 уровня доверительной вероятности: 1 уровень: β = 0,95 или β =95%; 2 уровень: α = 0,99 или β = 99%; 3 уровень: β = 0,999 или β = 99,9%. **В соответствии с (**4 **и** 4’**) всегда α + β = 1 или α% + β% = 100%,** т.е. по известному α определяется β и наоборот. В физической культуре и спорте достоверными считаются выводы: по первому уровню α и β - результат использования которых не допускает физические травмы; по второму уровню α и β - результат использования которых может привести к физическим травмам; по третьему уровню α и β - результат использования которых может привести к очень серьёзным нежелательным последствиям.  **Объектами изучения математической статистики** являются статистические совокупности. **Основная задача математической статистики** состоит в том, чтобы с помощью анализа выборочных совокупностей получить выводы относительно этих совокупностей или относительно их генеральных совокупностей с приведением уверенности в справедливости полученных выводов поуровню значимости или доверительной вероятности. Совокупности с количественными случайными данными могут различаться следующими основными статистическими характеристиками: **среднее арифметическое значение**, равное сумме всех её значений, делённой на их количество, т.е. х̅ = (х1+х2+х3+…+хп) / п = Σ xi / п (6); **отклонение i-го значения от среднего значения**: ( xi - х̅ ). **Главное свойство точного среднего значения совокупности** состоит в том, что точная сумма отклонений всех значений от него всегда равна нулю, т.е. Σ ( xi - х̅ ) = 0 (7). Это связано с тем, что сумма отрицательных отклонений данных от своего среднего всегда равна сумме положительных отклонений от него. Например, для совокупности xi: 15,17,11,9,12,7,14,11, где п = 8, х̅ =12 и Σ ( xi - х̅ ) = 3+5-1-3+0-5+2-1 = 10-10=0. Не смотря на то, что все отклонения от среднего по отдельности показывают разброс значений относительно своего среднего, нулевая сумма (7) не может характеризовать одну из главных особенностей совокупности – её общего разброса от среднего. Чем меньше этот разброс, тем совокупность плотнее, компактнее, однороднее.Поэтому, для получения показателя, характеризующего эту особенность, чтобы освободиться от минусов складывают квадраты отклонений от среднего, которые всегда положительные, и полученную сумму делят на число отклонений. В результате этих преобразований получается такой показатель в виде положительного числа, названного **дисперсией** (от лат. dispersio – рассеяние) и обозначаемого во многих учебниках по математической статистике малой буквой греческого алфавита сигма в квадрате, т.е. σ² = Σ( xi - х̅ )² / п (8).Т.к. единицы дисперсии квадратные по отношению к единицам данных, то чаще для характеристики разброса данных в совокупности используют величину, равную положительному значению корня квадратного из дисперсии, совпадающую с единицами данных и называемую **среднеквадратическим или стандартным отклонением**, т.е. σ = ( Σ( xi - х̅ )²/ п) ½ (9) - для больших совокупностей и σ = ( Σ( xi - х̅ )²/(п-1)) ½ (10) - для малых совокупностей. Чем меньше дисперсия совокупности , а значит и её стандартное отклонение, тем меньше разброс данных в ней, тем выше её компактность или однородность . В случае их равенства нулю, все статистические данные одинаковые, совпадают по значению со своим средним и превращаются из случайных данных в определённое (детерминированное) значение уровня рассматриваемого в совокупности статистического признака с учётом абсолютной погрешности измерения данных. Значения среднего арифметического х̅, дисперсии σ² и стандартного отклонения σ в программе Microsoft Excel определяются автоматически по введённым данным нажатием кнопки справа от значка Σ и выбором нужного действия из перечисленного списка. При этом, определение дисперсии σ² и стандартного отклонения σ выполняется в их нахождении после активации строки «другие функции». Т.к. совокупности с разными средними значениями могут иметь одинаковые стандартные отклонения, то для сравнения их компактности вводится относительный показатель разброса (изменения, вариации) данных около своего среднего – **коэффициент вариации** **v**. Он равен отношению стандартного отклонения совокупности к её среднему, т.е. v = σ/ х̅ (11) – безразмерная величина или v = σ\*100%/ х̅ (11’) - величина в процентах, которая может изменяться от 0 до +∞ в зависимости от величины данных. В математической статистике принято среднее значение приводить в виде **х̅± m х̅** (12), где m х̅ = σ /(n) ½ (13) его **ошибка репрезентативности** или просто – ошибка среднего. Имея числовое выражение (12) и значение объёма п по соответствующим формулам не сложно определить остальные, из выше рассмотренных, характеристики количественной совокупности. Значения всех вышеперечисленных характеристик можно определять автоматически с помощью компьютерной программы Microsoft Excel. В качестве примера определим их для совокупности xi: 15,17,11,9,12,7,14,11. Шаг1. Внесём данные в столбец А выбранного листа Excel. В результате видно, что п=8. Шаг 2. Активируем рабочую ячейку п+1=9. Нажимая на кнопку около значка Σ, можно выбрать любое из перечисленных в открывшемся меню действий. Выбираем «среднее». Нажимая на него и на Enter, получаем в рабочей ячейке х̅ =12. Здесь и далее найденное значение характеристики округляется до разумной точности. Например, округлять с одним запасным знаком по отношению к точности данных в совокупности. Эта цифра уже является сомнительной и нет смысла усложнять вычисления, приводя следующие за ней ещё более сомнительные цифры. Шаг4. Освободив рабочую ячейку, активируем её. Нажимая на кнопку около значка Σ, выбираем в меню «Максимум». Нажимая на него и на Enter, получаем в рабочей ячейке xmax = 17. Шаг5. Освободив рабочую ячейку, активируем её. Нажимая на кнопку около значка Σ, выбираем в меню «Минимум». Нажимая на него и на Enter, получаем в рабочей ячейке xmin = 7. Шаг6. Освободив рабочую ячейку, активируем её. Нажимая на кнопку около значка Σ и выбирая в меню «другие функции», открываем окно мастер функций. В строке категории выбираем «статистические», открываем эти функции и в предложенном меню выбираем «дисп.в». Активируя эту строку и нажав ок с Enter, получаем в рабочей ячейке значение σ² = 10,6. Шаг 7. Повторив действия шага 3, в меню категории «статистические» находим строку «стандоткл.в». Нажимая на неё и нажав ок, открываем рабочие строки «число1 и число2». Копируя данные совокупности в строку «число1» и нажимая ок, получаем в рабочей ячейке σ = 3,3. С помощью калькулятора находим: по (1) R =10; по (11) и(11’) v = 0,275 или v= 27,5%. По формуле (12) находим m х̅ =1,2 и представляем среднее в виде (13): х̅± m х̅ = 12±1,2. Информация о количественной совокупности в виде её значений: **п; xmax; xmin; х̅± m х̅** (14) является достаточной для использования многих методов математической статистики.