

ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ



- Основные понятия математической статистики, генеральные и выборочные совокупности, методы формирования выборочной совокупности
- Относительные величины, виды и принципы вычисления относительных величин.
- Вариационные ряды, средние величины, методика вычисления средних величин.
- Вариабельность, методика вычисления среднего квадратического отклонения и коэффициента вариации.
- Доверительный интервал, вероятность безошибочного прогноза.
- Репрезентативность, ошибка репрезентативности, доверительные границы средней арифметической.

СТАТИСТИКА

СТАТИСТИКА - это общественная наука, которая изучает количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной.

САНИТАРНАЯ (МЕДИЦИНСКАЯ) СТАТИСТИКА - это раздел статистики, изучающий состояние здоровья населения (показатели общественного здоровья) и деятельность лечебно-профилактических учреждений (ЛПУ), то есть состоит из статистики здоровья и статистики здравоохранения.

Предметами изучения медицинской статистики являются: здоровье населения в целом и отдельных возрастно-половых групп; выявление и установление зависимостей между уровнем здоровья и факторами окружающей среды; данные о сети, деятельности, кадрах учреждений здравоохранения; оценка статистической достоверности результатов медико-биологических, клинических и экспериментальных исследований.

Целью исследования в медицине и здравоохранении является выявление закономерностей изучаемого явления на основе проверки статистических (рабочих) гипотез, сформулированных в начале исследования. В зависимости от цели конкретизируются частные задачи и составляются план проведения всей работы, а также детальные программы сбора, разработки и статистического анализа собранного материала.

СТАТИСТИКА

РАБОЧЕЙ ГИПОТЕЗОЙ называется эмпирически не проверенное предположение, предсказывающее существование некоторой зависимости между переменными или объектами. Рабочие гипотезы используются для выработки предварительного плана научного исследования.

Статистическое исследование начинается с определения объекта наблюдения, единицы наблюдения и учетных признаков.

ОБЪЕКТОМ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ является статистическая совокупность, состоящая из единиц, о которых должны быть собраны статистические сведения, взятая в определенных границах пространства и времени. Например, при сборе данных о заболеваемости объектом наблюдения является совокупность всех заболеваний или всех больных среди населения на данной территории за определенный период времени.

ЕДИНИЦА СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ – это составная часть объекта наблюдения, подлежащая изучению и регистрации в соответствии с программой исследования. Например, отдельный случай заболевания или больное лицо, входящее в состав объекта наблюдения.

Учетные признаки

УЧЕТНЫЕ ПРИЗНАКИ - это медико-биологические характеристики, регистрируемые у единицы наблюдения в соответствии с целями и задачами исследования. Такими признаками могут быть: пол, возраст, место жительства, диагноз, дата заболевания, длительность болезни, ее исход и т.д.

ВИДЫ УЧЕТНЫХ ПРИЗНАКОВ, используемые в медицинской статистике:

- сходства, по которым единицы наблюдения объединяются в статистическую совокупность;
- различия, которые отличают единицы наблюдения между собой;
- факторные, которые влияют на изучаемое явление;
- результативные, которые изменяются под влиянием факторных признаков.

В зависимости от способа регистрации встречаются следующие ТИПЫ УЧЕТНЫХ ПРИЗНАКОВ:

- качественные (атрибутивные), которые могут быть выражены словесно, описательно;
- количественные, фиксирующие числовые значения признака.

Основные понятия математической статистики

СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ - это группа, состоящая из большого числа относительно однородных элементов (единиц наблюдения), взятых вместе в известных границах времени и пространства. Необходимо различать два основных вида статистических совокупностей - генеральную и выборочную.

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ состоит из всех возможных единиц наблюдения, которые могут быть к ней отнесены в соответствии с целью исследования. Генеральная совокупность в медико-биологических исследованиях используется редко. Большинство исследований выполняется с использованием выборочной совокупности.

ВЫБОРОЧНАЯ СОВОКУПНОСТЬ - это часть генеральной совокупности, отобранная специальным методом и предназначенная для характеристики генеральной совокупности.

Методы статистических наблюдений

- сплошное исследование, при котором изучаются все доступные единицы наблюдения генеральной совокупности;
- выборочное, при котором изучается определенная часть единиц наблюдения, наиболее полно характеризующие статистическую совокупность в целом.

В зависимости от продолжительности исследования применяются следующие виды статистических наблюдений.

Единовременное наблюдение – это исследование, при котором статистические данные собираются на определенный (критический) момент времени. Например, перепись населения.

Текущее наблюдение – это непрерывное, повседневное исследование, производимое в течение определенного периода: месяца, полугода, года. Например, изучение уровня заболеваемости населения города или региона.

Этапы исследования



1 ЭТАП - составление программы и плана исследования;

2 ЭТАП - статистическое наблюдение;

3 ЭТАП - группировка и разработка статистического материала;


4 ЭТАП - анализ результатов исследования.

1 этап

ПРОГРАММА ИССЛЕДОВАНИЯ включает вопросы, что и в каком направлении изучать, с обозначением объекта и единиц наблюдения, учетных признаков, методов сбора (анкеты, бланки, первичные документы), разработки и анализа материала (макеты таблиц, выбранные статистические критерии).

ПЛАН ИССЛЕДОВАНИЯ отвечает на вопросы: где, сколько, когда, кто и как выполняет исследование, и включает:

- выбор места проведения исследования;
- пути формирования объекта наблюдения (объем выборки, время, способы сбора материала);
- определение единицы наблюдения;
- способы разработки материала;
- сроки работ по этапам;
- выбор исполнителей;
- финансирование исследования;
- инструкции исполнителям, организационное и методическое сопровождение.



ПРОГРАММА СБОРА МАТЕРИАЛА представляет собой первичный учетный документ (бланк, карта, анкета), в который включены учетные признаки, соответствующие цели исследования и подлежащие регистрации.

При составлении учетного документа необходимо соблюдать следующие правила:

1) документ должен иметь четкое заглавие, в котором сформулирована единица наблюдения;

2) учетные признаки должны быть указаны краткими названиями и соответствовать цели исследования;

3) на каждый вопрос исследования следует предусмотреть варианты ответов в соответствии с выделенными группами единиц наблюдения.

Программа разработки материала предусматривает определение критериев группировки единиц наблюдения и составление макетов статистических таблиц.

3 этап

КОНТРОЛЬ - это проверка собранного материала с целью отбора учетных документов, имеющих дефекты, для их последующего исправления, дополнения или исключения из исследования.

ШИФРОВКА - это применение условных обозначений изучаемых признаков с последующим назначением шифра каждой единице наблюдения. При ручной обработке материала шифры могут быть цифровые, буквенные, знаковые; при машинной, как правило, цифровые.

ГРУППИРОВКА - это распределение единиц наблюдения на однородные группы по характеру или величине изучаемых признаков.

ВИДЫ ГРУППИРОВОК:

- типологическая - это группировка по атрибутивным (качественным) признакам. Например: пол, профессия.

- вариационная - это группировка на основе признаков, имеющих числовое выражение. Например: возраст, стаж.

СВОДКА ДАННЫХ УЧЕТА - занесение полученных после подсчета цифровых данных в таблицы.

Для дальнейшего анализа материала необходимо произвести РАСЧЕТЫ СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ и средних величин в соответствии с программой исследования и выполнить ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ.

4 этап



- интерпретацию полученных различных статистических величин и графических изображений на основе сопоставления с нормативами, со средними уровнями аналогичных величин, со стандартами, с данными по другим учреждениям и территориям, литературными данными, в динамике;
- обобщение результатов исследования;
- литературное оформление работы;
- выявление закономерностей;
- выводы;
- предложения для внедрения в практику;
- прогноз, рекомендации.

Относительные величины

Абсолютные величины, полученные непосредственно при измерении учетных признаков каждой единицы наблюдения, несут важную информацию о размере того или иного явления (количество родившихся, умерших; число коек в каждой больнице города, число дней болезни каждого больного и др.). Они служат основой для вычисления относительных величин.

Относительные величины вычисляются путем отношения (деления) одной абсолютной величины на другую с последующим умножением полученной дроби на требуемое основание (100, 1000, 10 000, 100 000 и т.д.).

Соответственно, относительные показатели могут быть выражены в процентах (%), промилле (‰), продецимилле (‱), просантимилле (‱). Для их условного обозначения применяется знак “*P*”.

Виды относительных показателей

ИНТЕНСИВНЫЕ ПОКАЗАТЕЛИ отражают ЧАСТОТУ (встречаемость, уровень) явления в СРЕДЕ, ПОРОДИВШЕЙ это явление и, как правило, вычисляются на основании 1000 (в промилле, ‰). Если полученное значение меньше единицы, целесообразно использование множителей 10 000 (‰), 100 000 (‱).

$$P_{\text{Интенс.}} = \frac{\text{Абсолютный размер явления} * 1000}{\text{Абсолютный размер среды}}$$

ЭКСТЕНСИВНЫЕ ПОКАЗАТЕЛИ отражают структурный состав изучаемой совокупности. Они характеризуют отношение части статистической совокупности к совокупности в целом (долю, удельный вес, часть от целого), т.е. отношение отдельного элемента к итогу. Их нельзя использовать для определения уровня изучаемого явления, они демонстрируют только соотношение его частей и чаще выражаются в %.

$$P_{\text{Экст.}} = \frac{\text{Абсолютный размер части явления} * 100}{\text{Абсолютный размер явления в целом}}$$

Виды относительных показателей

ПОКАЗАТЕЛИ СООТНОШЕНИЯ применяются, когда необходимо определить взаимоотношение несвязанных между собой совокупностей. Например, обеспеченность населения больничными койками или врачами, соотношение средних медработников и врачей и др. Они вычисляются как отношение величины одного явления к другому явлению и выражаются, в основном, в продецимилле, реже - в процентах, промилле и др.

$$P_{\text{Соотн.}} = \frac{\text{Абсолютный размер одного явления} * 10000}{\text{Абсолютный размер второго явления}}$$

Для анализа изменения изучаемого явления во времени вычисляются ДИНАМИЧЕСКИЕ ПОКАЗАТЕЛИ на основе динамического ряда.

Динамические ряды

Динамическим называется ряд чисел, состоящий из однородных сопоставимых величин, характеризующих изменения какого-либо явления за определенные отрезки времени.

Виды динамических рядов:

- простой (состоит из абсолютных величин);
- сложный (из относительных и средних);
- моментный (состоит из величин, характеризующих размеры явлений на определенные даты);
- интервальный (состоит из величин, характеризующих размеры явления за определенный интервал времени).

Числа динамического ряда называются УРОВНЕМ.

Показатели динамического ряда

- АБСОЛЮТНЫЙ ПРИРОСТ - разность уровней изучаемого явления в данном и предыдущем периоде, может быть положительным, нулевым или отрицательным (убыль);

- ПОКАЗАТЕЛЬ НАГЛЯДНОСТИ - отношение каждого последующего уровня к исходному, принятому за 100%;

- ПОКАЗАТЕЛЬ РОСТА, или темп роста - отношение каждого последующего уровня к предыдущему уровню, принятому за 100%;

- ПОКАЗАТЕЛЬ ПРИРОСТА, или темп прироста - отношение абсолютного прироста к предыдущему уровню, принятому за 100%. Как и абсолютный прирост этот показатель может выражаться положительным (прирост), нулевым (стабильность) или отрицательным (убыль) значением.

Графическое представление

ИНТЕНСИВНЫЕ показатели, а также показатели СООТНОШЕНИЯ могут быть наглядно представлены в виде 4-х основных типов диаграмм: столбиковые, линейные, картограммы и картодиаграммы. Картограмма - это географическая карта с различной штриховкой. Картодиаграмма - это географическая карта с нанесенными на нее диаграммами.

ЭКСТЕНСИВНЫЕ показатели графически могут быть изображены секторной или внутрестолбиковой диаграммой.

Для графического изображения динамических показателей наиболее часто применяются линейная и столбиковая диаграммы. Радиальная (лепестковая) диаграмма является частным видом линейной диаграммы, построенной на полярных координатах. Она используется при необходимости изобразить графически динамику явления за замкнутый цикл времени.

Вариационный ряд

Типы переменных:

Непрерывные переменные (continuous variables) могут принимать любые численные значения, которые естественным образом упорядочены на числовой оси (например, рост, вес, АД, РОЭ).

Дискретные переменные (discrete variables) могут принимать счётное множество упорядоченных значений, которые могут просто обозначать целочисленные данные или ранжировать данные по степени проявления на упорядоченной ранговой шкале (например, клиническая стадия опухоли, тяжесть состояния пациента).

Категориальные переменные (categorical variables) являются неупорядоченными и используются для качественной классификации (пол, цвет глаз, место жительства); в частности, они могут быть бинарными (дихотомическими) и иметь категорические значения: 1/0, да/нет, имеется/отсутствует.

Вариационный ряд

ВАРИАЦИОННЫЙ РЯД – это однородная в качественном отношении статистическая совокупность, отдельные единицы которой характеризуют количественные различия изучаемого признака или явления.

Количественные (числовые) данные предполагают, что переменная принимает некоторое числовое значение. Из них выделяют дискретные данные и непрерывные. Также и вариационные ряды делятся на *дискретные* и *непрерывные*.

Численное значение каждого отдельного признака или явления, входящего в вариационный ряд, называется вариантой и обозначается буквой V или x , или y .

Виды вариационных рядов

Вариационный ряд, в котором каждая варианта встречается только один раз, называется ПРОСТЫМ.

При увеличении числа наблюдений, как правило, встречаются повторяющиеся значения вариант. В этом случае создается СГРУППИРОВАННЫЙ ВАРИАЦИОННЫЙ РЯД, где указывается число повторений (частота, обозначается буквой « p »).

РАНЖИРОВАННЫЙ ВАРИАЦИОННЫЙ РЯД состоит из вариант, расположенных в порядке возрастания или убывания.

ИНТЕРВАЛЬНЫЙ ВАРИАЦИОННЫЙ РЯД составляют с целью упрощения последующих вычислений, выполняемых без использования компьютера, при очень большом числе единиц наблюдения (более 1000).

НЕПРЕРЫВНЫЙ ВАРИАЦИОННЫЙ РЯД включает значения вариант, которые могут выражаться любыми значениями.

Если в вариационном ряде значения признака (варианты) заданы в виде отдельных конкретных чисел, то такой ряд называют ДИСКРЕТНЫМ.

Средние величины

Средняя величина характеризует общий количественный уровень изучаемого признака и является групповым свойством статистической совокупности. Она нивелирует, ослабляет случайные отклонения индивидуальных наблюдений в ту или иную сторону и выдвигает на первый план основное, типичное свойство изучаемого признака.

В медицине средние величины широко используются:

Для оценки состояния здоровья населения: характеристики физического развития (рост, вес, окружность грудной клетки и пр.), ***выявления распространенности и длительности различных заболеваний, анализа демографических показателей*** (естественного движения населения, средней продолжительности предстоящей жизни, воспроизводства населения, средней численности населения и др.).

Для изучения деятельности ЛПУ, медицинских кадров и оценки качества их работы, планирования и определения потребности населения в различных видах медицинской помощи (среднее число обращений или посещений на одного жителя в год, средняя длительность пребывания больного в стационаре, средняя продолжительность обследования больного, средняя обеспеченность врачами, койками и пр.).

Средняя величина

Для характеристики санитарно-эпидемиологического состояния (средняя запыленность воздуха в цехе, средняя площадь на одного человека, средние нормы потребления белков, жиров и углеводов и т. д.).

Для определения медико-физиологических показателей в норме и патологии, при обработке лабораторных данных, для установления достоверности результатов выборочного исследования в социально-гигиенических, клинических, экспериментальных исследованиях.

Средняя величина

Вычисление средних величин выполняется на основе вариационных рядов.

Общими характеристиками значений признака, отражаемого в вариационном ряду, являются средние величины. Среди них наиболее применяемые: **средняя арифметическая величина M (среднее выборочное), мода M_o и медиана M_e .**

Наиболее типичной характеристикой значений вариант является средняя арифметическая величина (M). В математической статистике она обозначается \bar{X} .

МОДОЙ (M_o) называют значение наиболее часто встречающейся варианты.

МЕДИАНА (M_e) – это значение варианты, делящей ранжированный вариационный ряд пополам (с каждой стороны медианы находится половина вариант). В редких случаях, когда имеется симметричный вариационный ряд, значения моды и медианы совпадают со значением средней арифметической.

Средняя арифметическая величина

СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ ВЕЛИЧИНА (M, \bar{X}) – это общая количественная характеристика определенного признака изучаемых явлений, составляющих качественно однородную статистическую совокупность. Различают среднюю арифметическую простую и взвешенную.

Вычисление простой средней арифметической проводится по формуле:

$$M = \frac{\sum V}{n},$$

где: M - средняя арифметическая простая;

$\sum V$ - сумма вариантов;

n - число наблюдений.

В сгруппированном вариационном ряду определяют взвешенную среднюю арифметическую. Формула для ее вычисления:

$$M = \frac{\sum Vp}{n},$$

где: M - средняя арифметическая взвешенная;

$\sum Vp$ - сумма произведений вариант на их частоты;

n - число наблюдений.

Параметры вариабельности признака

Размах;

Амплитуда;

Среднее квадратическое отклонение;

Коэффициент вариации.

Размах указывает на максимальную (V_{\max}) и минимальную (V_{\min}) варианты в ряду.

Амплитуда (A_m) является разностью этих вариантов: $A_m = V_{\max} - V_{\min}$.

Дисперсия вариационного ряда

Основной, общепринятой мерой колеблемости вариационного ряда являются ДИСПЕРСИЯ (D). Но наиболее часто применяется более удобный параметр, вычисляемый на основе дисперсии - среднее квадратическое отклонение (σ). Оно учитывает величину отклонения (d) каждой варианты вариационного ряда от его средней арифметической ($d = V - M$).

Дисперсия вариационного ряда является средним квадратом отклонения вариантов от средней арифметической и вычисляется по формуле

$$D = \frac{\sum d^2}{n}$$

Среднее квадратическое отклонение

Обозначени «Сигма» (σ)

е:

Среднее квадратическое отклонение определяют по формуле:

$$\sigma = \sqrt{\frac{\sum d^2}{n}}.$$

Указанная формула применяется при числе наблюдений (n) больше 30. При меньшем числе n значение среднего квадратического отклонения будет иметь погрешность, связанную с математическим смещением ($n - 1$). В связи с этим, более точный результат может быть получен с помощью учета такого смещения в формуле расчета стандартного отклонения:

$$s = \sqrt{\frac{\sum d^2}{n-1}}.$$

Стандартное отклонение

СТАНДАРТНОЕ ОТКЛОНЕНИЕ (s) – это оценка среднеквадратического отклонения случайной величины X относительно её математического ожидания на основе несмещённой оценки её дисперсии.

При значениях $n > 30$ среднее квадратическое отклонение (σ) и стандартное отклонение (s) будут одинаковыми ($\sigma = s$). Поэтому в большинстве практических пособий эти критерии рассматриваются как синонимичные. В программе Excel вычисление стандартного отклонения может быть выполнено функцией =СТАНДОТКЛОН(диапазон). А с целью расчета среднего квадратического отклонения требуется создать соответствующую формулу.

Среднее квадратическое или стандартное отклонение позволяет определить, насколько значения признака могут отличаться от среднего значения.

Коэффициент вариации

Сравнение variability признаков различной размерности может быть выполнено с помощью КОЭФФИЦИЕНТА ВАРИАЦИИ. Он выражает разнообразие в процентах от средней величины, что позволяет производить сравнение различных признаков. Коэффициент вариации в медицинской литературе обозначается знаком «*C*», а в математической «*v*» и вычисляется по формуле:

$$C = \frac{\sigma}{M} \times 100\% .$$

Значения коэффициента вариации менее 10% свидетельствует о малом рассеянии, от 10 до 20% – о среднем, более 20% – о сильном рассеянии вариант вокруг средней арифметической.


Ошибка репрезентативности

Величина ошибки репрезентативности зависит как от объема выборки, так и от вариабельности признака. Чем больше число наблюдений, тем ближе выборка к генеральной совокупности и тем меньше ошибка. Чем более изменчив признак, тем больше величина статистической ошибки.

На практике для определения ошибки репрезентативности в вариационных рядах пользуются следующей формулой:

$$m = \frac{\sigma}{\sqrt{n-1}},$$

где: m – ошибка репрезентативности;
 σ – среднее квадратическое отклонение;
 n – число наблюдений в выборке.


$$m = \sqrt{\frac{Pq}{n-1}},$$

где: P – величина относительного показателя, выраженного в процентах, промилле и т.д.;

q – величина, обратная P и выраженная как $(1-P)$, $(100-P)$, $(1000-P)$ и т. д., в зависимости от основания, на которое рассчитан показатель;

n – число наблюдений в выборочной совокупности.