

ЯГМА
Кафедра медицинской физики
Лечебный факультет

1 курс
1 семестр

«Элементы математической статистики»

Составил: Дигурова И.И.

2004 г.

1. Математическая статистика. Ее виды, особенности, задачи.

Математическая статистика – раздел математики, посвященный математическим методам обработки, систематизации и использования статических данных для практических и научных целей.

Задачей этого раздела является разработка практических методов, регистрации, описания, анализ экспериментальных данных, получаемых в опытах с массовыми явлениями.

Особенностью статистики является изучение массовых, случайных явлений в условной неопределенности. Достоверность выводов зависит от числа объектов исследования. На основе анализов и прогнозов вырабатывается оптимальное решение.

Статистика подразделяется на:

- теоретическую (вырабатывает методы)
- прикладную (общая, отраслевая (экономическая статистика, метеорологическая, медицинская))

Медицинская статистика:

- ст. рождаемости
- ст. заболеваемости
- ст. смертности
- ст. медицинских учреждений

Биологическая статистика (=биометрия) – включает статистические методы, используемые в различных биологических исследованиях (в цистологии, микробиологии).

Статистика:

- описательная (комплекс методов сбора, группировки данных и представления их в виде таблиц, графиков...)
- аналитическая (делает заключения, выводы с целью практического применения)

2. Основные понятия описательной статистики.

Их характеристика

1. Генеральная совокупность – подлежащая изучению совокупность однородных элементов, которая характеризуется некоторым признаком. Например, нас интересует распространенность данного заболевания в определенном регионе, тогда генеральная совокупность, это все население региона. Если необходимо выразить мужчин и женщин отдельно по этому заболеванию, то получаем 2 генеральные совокупности.

Количество объектов, входящих в генеральную совокупность называется объемом генеральной совокупности (N)

Генеральная совокупность можно изучать по некоторой ее части.

2. Выборочная совокупность - часть генеральной совокупности, выбираемая для статистической обработки (выборка) (объем выборки -n). Свойства объектов выборки должны соответствовать свойствам генеральной совокупности.

Результаты исследования некоторого признака генеральной совокупности, будут более надежны, если выборку образовывать случайным образом. Элементы выборки берутся наугад. Каждый объект может попасть в выборку с одинаковой вероятностью. Главным вопросом является: как определить объем выборки, необходимой для получения необходимого результата.

3. Варианта – значение признака для каждого элемента выборки (x)

Признаки могут быть качественными и количественными

Количественные делятся на непрерывные (масса тела) и дискретные (количество волос)

Признак, имеющий значение от одного объекта к другому называется варьирующимся. Если количественный признак лежит в интервале – интервальный.

4. Частота – количество объектов с конкретным числовым значением признака

x_i	35	36	37	38	39	40	41
n_i (p_i):	2	4	5	6	7	7	2

5. Частность или относительная частота – доля варианта с данным значением признака (n_i/n)

3. Ряды распределения и способы их представления.

Ряд распределения – это последовательность качественных-количественных значений признака и частоты его встречаемости.

Ряд, составленного на основе качественного признака – атрибутивных количественного – вариационный.

Рассмотрим подробнее распределение количественного признака. Значение признака, записанное для всех элементов выборки в том порядке, в каком они были получены образуют простой (упорядоченный) статистический ряд.

1 2 3 4 5 6
170 165 171 165 163 174

Из данных видно: некоторые значения вариант повторяются. Для сокращения записи данные располагаются в упорядоченном виде с указанием частот. Такой ряд называется упорядоченным (=ранжированным).

n_i 1 2 1 1 1 1

x_i 163 165 160 171 174

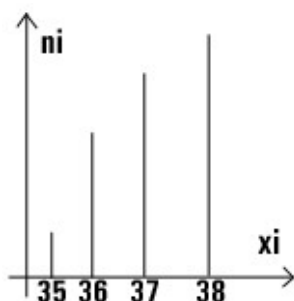
Вариационные ряды могут быть непрерывными и дискретными

Способы представления рядов:

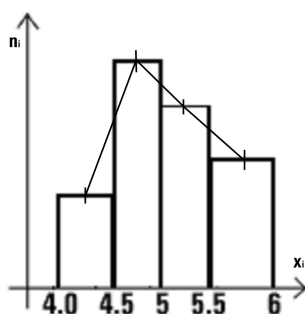
1. Табличный (см. выше)
2. Аналитический (с помощью формул)
3. Графический (строится на основании табличных данных)

Способы графического представления:

А) диаграмма в отрезках – совокупность вертикальных прямых /отрезков. Способ удобен для представления дискретных признаков при небольшом объеме совокупности.

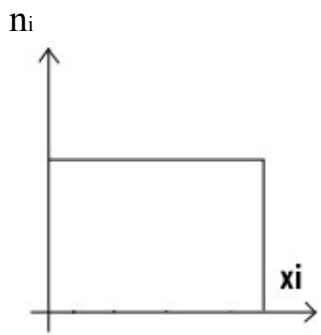


Б) гистограмма – совокупность прилегающих друг к другу прямоугольников. Способ используется для изображения интервального ряда. На оси X откладываются интервалы значения варианта. На каждом из них (на основании) строят прямоугольник. Его высота зависит от частоты встречаемости данной величины.

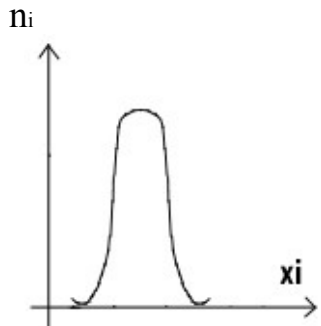


В) полигон частот – ломаная линия, соединяющая точки, являющиеся серединами интервалов.

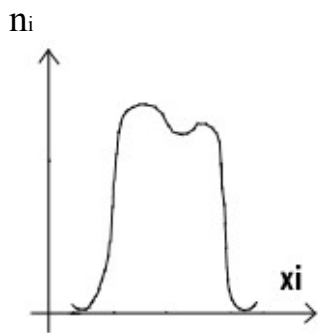
Г) Вариационные кривые в зависимости от значения распределения.



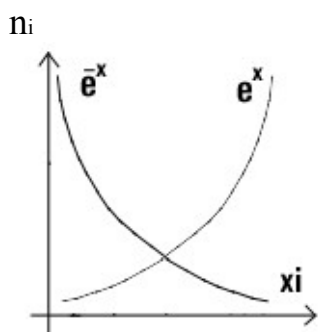
Прямоугольное распределение объема совокупности



Колоколообразное (= унимодальное)



Бимодальное



Экспоненциальное

1. Меры положения частного распределения, их характеристика.

На практике ряды распределения описываются различными числовыми характеристиками (мерами).

1 Мода (M_o) – это варианта, наиболее часто встречающаяся в совокупности (= модальное значение).

2 Медиана (M_e) – это величина, делящая ранжированный ряд на 2 равные части. Так же она делит площадь под кривой распределения. Для того, чтобы определить M_e надо ранжировать ряд (в порядке возрастания), вычислить номер, под которым стоит медиана.

$N/2$ – Для четных, $(N+1)/2$ – Для нечетного количества объектов

3 Средняя арифметическая простая – это частное деление суммы всех значений признака на их общее число объектов

$$\bar{X} = (X_1 + X_2 + X_3 \dots + X_n) / N$$

Сумма всех $\langle + \rangle$ и $\langle - \rangle$ отклонений от x равно «0».

Среднюю арифметическую простую вычисляют для неупорядоченных рядов в тех случаях, когда каждая варианта встречается 1 раз.

4 Средняя взвешенная

Если в совокупности отдельные варианты встречаются неоднократно, то вычисляется средняя взвешенная – это величина, полученная суммированием произведений числовых значений вариант на их частоты с последующим делением суммы на количество всех вариант.

$$\bar{X} = (x_1 n_1 + x_2 n_2 + \dots + x_n n_n) / N \quad x_1 n_1 + y_2 n_2$$

5 Средняя квадратическая используется, если признаки выражаются мерами площади. Пример: размер колонии микробов, листовых пластинок.

$$\bar{X}_q = \sqrt{\frac{\sum_{i=1}^{n_1} n_1 x_2^2}{N}}$$

6 Средняя гармоническая, кубическая, геометрическая

2. Меры рассеяния частного распределения. Его характеристика.

Разброс числовых значений вариант (генеральной, выборочной совокупности) относительно средних значений характеризуется мерами рассеяния.

1. Лимит – минимальная и максимальная варианта совокупности. (Xmin, Xmax)

2. Вариационный размах – разность между максимальным и минимальным значением $R = X_{\max} - X_{\min}$

3. Индивидуальное отклонение – разность между числовым значением варианты и средним арифметическим всей совокупности $D_i = X_i - \bar{X}$

4. Дисперсия – мера рассеяния, полученная суммированием квадратов индивидуальных отклонений и последующим делением суммы на объем совокупности.

$$\sigma^2 = \frac{\sum d_i^2}{N}$$

σ^2 - для генеральной

s^2 - для выборочной совокупности

Если число объектов менее 30, то рассчитывается исправленная дисперсия (Сигма с крышей)

$$\hat{\sigma}^2 = \frac{\sum d_i^2}{N-1}$$

Где N-1 – число степеней свободы. Это число на 1 меньше, чем весь объем свободности

5. Стандартное (среднее квадратичное) отклонение. Эта мера рассеяния, равная корню квадрата дисперсии, $S = \sqrt{S^2}$. Чем сильнее варьирует признак, тем больше величина среднего квадратного отклонения.

6. Коэффициент вариации - мера рассеяния равна, отношению стандартного отклонения к средней арифметической $V = (S/X)100\%$

При нормальном распределении коэффициент вариации не $> 50\%$, а часто гораздо ниже (приблизительно 20%)

3. Выборочный метод. Выборки, их виды и требования к ним.

Для того, чтобы получить исчерпывающую информацию о состоянии генеральной совокупности нужно учесть весь ее состав без исключения. Но не всегда есть возможность или необходимость прибегать к сплошному исследованию. В целях экономии времени и средств, анализу подвергается часть совокупности выборки, по ней судят о состоянии всей совокупности в целом.

Если число объектов менее 30, то выборка называется малой. В зависимости от способов формирования, выборки бывают повторные – с возвратом, неповторные – без возврата

Требования к выборкам

А) Рендомизация - каждая варианта генеральной совокупности имеет одинаковую вероятность для попадания в выборку.

Репрезентативность – состав и структура выборки должны соответствовать составу и структуре генеральной совокупности.

	Генеральная совокупность	Выборочная
Объем	N	n
Среднее значение	M	\bar{X}
Дисперсия	σ^2	S в квадрате
Стандартное отклонение	σ	S

4. Ошибки репрезентативности, их особенности.

Возникающие отклонения выборочных показателей от параметров генеральной совокупности называются – ошибками репрезентативности. Параметрами называются характеристики, относящиеся к генеральной

совокупности. Характеристики, относящиеся к выборке называются оценками параметров. Ошибки репрезентативности бывают случайными и систематическими, устранимыми и неустраняемым. Устраняемая ошибка предотвращается правильной организацией исследования и четким ведение протокола.

Неустраняемые ошибки заложены в природе статистических методов. Фактически они являются ошибками репрезентативности. Это своеобразные показатели вариаций выборочных характеристик по отношению к таким же характеристикам генеральной совокупности. Величина ошибки зависит от объема выборки, степени вариации признаков, способа отбора вариантов. При увеличении числа вариантов выборки ошибки– 0.

Ошибка (M_x)

$M_x = (S) / \sqrt{n}$

$M_x = (S) / \sqrt{2n}$

$M_x = (V) / \sqrt{2n}$

V-коэффициент вариации

Показатель точности оценки параметров.

Чтобы получить определенное представление о точности, с которой определяется тот или иной средний результат, принято использовать показатели точности.

$C = (\bar{M}_x / \bar{X}) 100\%$; Если известно значение коэффициента вариации, то используется $C = V / \sqrt{n}$

Точность достаточная, если $C = 3-5\%$

5. Нормальное распределение, его виды, формулы, графики, особенности.

Для того, чтобы оценить закон распределения переменной случайной величины, нужно найти функциональную зависимость между числовыми значениями, которые она может принимать и вероятностью этих значений. В пределах заданного интервала непрерываемая случайная величина может принимать любые числовые значения. Речь идет о значениях, которые она может принимать с определенной вероятностью.

Виды нормального распределения

А) Эмпирическое – получается опытным путем на основе статистического исследования. В этом случае объем совокупности всегда конечен

Б) Теоретическое – абстрактная математическая модель. Ее используют для сравнения и оценки опытных распределений по разным статистическим критериям.

В) Стандартное – оно используется в качестве стандарта при оценке любых данных.

Г) Общее – нормальное распределение – оно описывается формулой Гаусса – Лапласа, которая выражает зависимость между вероятностью и нормированным отклонением

$$y = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

y – значение функции – частота встречаемости вариант с данным значением признака

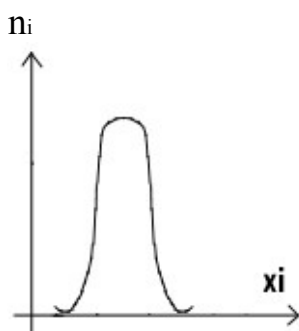
$$t = \frac{x_i - \bar{x}}{\sigma}$$

Сигма – стандартное отклонение, t – нормированное отклонение.

Нормированное отклонение определяет отклонение значение варианты от среднего арифметического (N) выраженного в долях от стандартного отклонения

График общего нормального распределения

Его строят в прямой системе координат. По горизонтали – числовые значения признака (возрастание), по вертикали – значение Y , соответствует числу данных вариант.



Особенности нормального распределения

1 Кривая симметрична относительно среднего арифметического.

2 Для строго нормального теоретического распределения средняя арифметическая мода и медиана совпадают

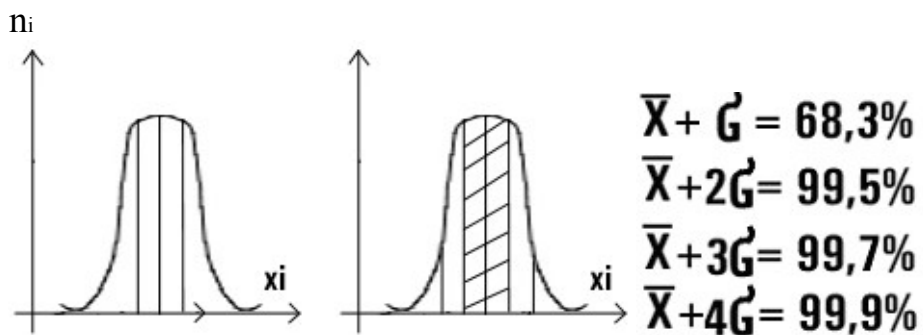
3 Положение кривой на горизонтальной оси определяется числовым значением средней арифметической

4 Вытянутость кривой зависит от стандартного отклонения. Чем выше сигма, тем упрощеннее график

Значение площади под кривой зависит от объема совокупности

5 Значение Y является наибольшим для средней арифметической.

6 Для каждой точки горизонтальной оси высота кривой определяется значением ординаты Y . Чем больше отклоняются числовые значения вариант от среднего арифметического, тем реже такие варианты встречаются в совокупности.



Площадь по кривой соответствует 100% всех наблюдений