

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ КАЗАНСКИЙ
ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ

КАФЕДРА СОЦИАЛЬНОЙ МЕДИЦИНЫ И ОРГАНИЗАЦИИ
ЗДРАВООХРАНЕНИЯ

А.Н. Галиуллин, С.Р. Камалова, А.И. Глушаков

ВВЕДЕНИЕ В МЕДИКО-СТАТИСТИЧЕСКИЙ АНАЛИЗ
ДАНЫХ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ

Учебное пособие

Казань 2000

ББК 51.1(2) УДК61.4.1:616-071(072)

Печатается по решению Центрального координационно-методического совета
Казанского государственного медицинского университета

Составители:

зав. кафедрой СмиОЗ, д.м.н., профессор Галиуллин А.Н.,
асс. кафедры СмиОЗ, к.м.н., Камалова С.Р.,
нач. отд. статистики Республиканского центра медико-социальных
исследований и экономики здравоохранения, к.м.н., Глушаков А.И.

Рецензенты:

доктор медицинских наук, профессор Вахитов Ш.М. доктор

медицинских наук, профессор Хамитова Р.Я.

Введение в медико-статистический анализ данных клинических исследований Учеб. пособие / Галиуллин А.Н., Камалова С.Р., Глушаков А.И. -Казань: КГМУ, 2000. - 14 с.

Цель учебного пособия - ознакомить практических врачей и студентов медицинских вузов с содержательно-формальными требованиями к подготовке данных клинических исследований для медико-статистического анализа. Изложены принципы стандартизации признаков, способы восстановления пропущенных значений и оценки «выскакивающих» наблюдений. Рассматриваются вопросы учета характера распределения, подбора методов для оценки связей и анализа их результатов. Изложены особенности многомерных оценок (шкалирования факторного анализа, главных компонент, кластер-анализа) для обнаружения информативной совокупности показателей, вопросы применения регрессионного и дисперсного анализа, методов теории распознавания образов для классификации различных состояний.

Казанский государственный медицинский университет, 2000

ОГЛАВЛЕНИЕ

Содержательный анализ, коррекция, формализация	4
Подготовка к проведению анализа	4
Классификация признаков .	5
Коррекция исходных данных	6
Формализация	6
Работа с пропущенными и неполными данными	7
Предварительная оценка и выбор статистических методов	8
Анализ связей	9
Подбор информативных показателей	10
Различные методы анализа данных	11
Литература	14

В практике анализа полученных во время научных и эпидемиологических исследований результатов необходимым условием является, как правило, оценка их достоверности и подтверждение наблюдаемого сходства (различия) с аналогичными данными или контрольными группами или решение задачи классификации в целях диагностики, прогнозирования исхода при различной лечебной тактике. В последние годы для этого широко используются отдельные программы или пакеты прикладных программ, реализованные на персональных компьютерах. Однако применение того или иного метода обработки данных является фактически заключительным этапом анализа результатов. Крайне важны четкая постановка задачи и подготовка материала для статистической обработки.

СОДЕРЖАТЕЛЬНЫЙ АНАЛИЗ, КОРРЕКЦИЯ, ФОРМАЛИЗАЦИЯ

В зависимости от поставленной задачи, объема и характера материала, вида данных и

их связей делается выбор методов статистической обработки как на этапах предварительного (для опенки характера распределения в исследуемой выборке), так и окончательного анализа, в соответствии с целями исследования.

Подготовка к проведению анализа

Крайне важным аспектом является проверка однородности выбранных групп наблюдения, в том числе контрольных, что может быть проведено или экспертным путем, или методами многомерной статистики (например, с помощью кластерного анализа). Но сначала необходимо составление вопросника (или тезауруса), в котором предусматривается стандартизованное описание признаков, в особенности при проведении эпидемиологических исследований, где необходимо единство в понимании и описании разными врачами одних и тех же симптомов, включая учет диапазонов (степени выраженности) их изменений

В случае существенности различий в регистрации исходных данных (субъективная оценка характера патологических проявлений различными специалистами) и невозможности их приведения к единому виду на этапе сбора информации, может быть осуществлена, так называемая, коррекция ковариант, которая предполагает нормализацию переменных, т.е. устранение ненормальностей показателей в матрице данных. Согласование мнений осуществляется с учетом специальности и опыта врачей, что позволяет затем сравнивать полученные ими результаты обследования между собой. Для этого могут использоваться многомерный дисперсионный и регрессионный анализы.

Классификация признаков

Признаки могут быть как однотипными, что бывает редко, так и разнотипными.

Под этим понимается их различная метрологическая оценка:

- количественные, или числовые признаки, — это замеренные в определенной шкале и в шкалах интервалов и отношений (I группа признаков);
- качественные, ранговые или балльные используются для выражения медицинских терминов и понятий, не имеющих цифровых значений (например, тяжесть состояния), и замеряются в шкале порядка (II группа признаков);
- классификационные, или номинальные (например, профессия, группа крови), - это замеренные в шкале наименований (III группа признаков).

Квантование количественных признаков может осуществляться как непосредственно исследователем, так и с помощью ПК при использовании специальной программы, учитывающей нормальные значения показателя;

различные граничные условия (незначительные, средние, высокие и крайне высокие значения), что предполагает введение макета квантования -квантометрическая шкала.

Во многих случаях делается попытка анализа крайне большого числа признаков, что, по мнению исследователей, должно способствовать повышению информативности представлений выборки. Однако выбор полезной информации, т.е. осуществление отбора признаков, является операцией совершенно необходимой, поскольку для решения любой классификационной задачи должны быть отобраны сведения, несущие не «шум» и не irrelevantную (не относящуюся к цели исследования), а полезную для данной задачи информацию. В случае, если это не осуществлено по каким-то причинам, исследователем самостоятельно или отсутствуют достаточно обоснованные критерии для снижения размерности пространства признаков по содержательным соображениям, борьба с избыточностью информации осуществляется уже формальными методами путем оценки информативности.

В отношении происхождения изменчивости данных следует иметь в виду, что существуют три основные ее группы:

- 1) систематическая, которая вызвана изучаемыми воздействиями, что и является обычно предметом исследования;

- 2) систематическая, связанная с условиями опыта (методами исследования), т.е. постоянными погрешностями (ошибками);
- 3) случайная (остаточная), вызванная нерегулярными изменениями в процессе исследования (Любищев А., 1986).

Коррекция исходных данных

Принципиально важным аспектом является анализ резких отклонений в исходных данных (аномальные или выделяющиеся выбросы), которые могут быть как следствием случайных колебаний, обусловленных не всегда известной природой генеральной совокупности, так и являться артефактом вследствие погрешности исследования (записи). В первом случае они требуют учета при подборе статистических методов для заключительной обработки данных, тогда как во втором такие наблюдения подлежат исключению по решению исследователя или путем использования формальных методов.

Статистические процедуры поиска резко выделяющихся случаев основаны на предположении однородности данных, в то время как «выскакивающие» наблюдения рассматриваются как атипичные, далеко удаляющиеся от центра распределения. Предложен ряд аналитических процедур для идентификации таких выбросов и оценки значимости их отклонения (Айвазян С. и соавт., 1983). Сомнительные наблюдения или полностью исключаются из дальнейшего рассмотрения или их вклад уменьшается с помощью весовой функции, убывающей по мере роста степени аномальности наблюдений.

Формализация

После определения перечня признаков для статистической обработки наступает этап формализации документа, где каждому значению признака (или симптому) соответствует определенное кодовое число.

Кодирование, как отмечают Н.М. Амосов и соавт. (1975), не должно изменять семантическую силу и смысл исходных данных, т.е. медицинское содержание информации.

Представление информации в квантованной форме обеспечивает фиксацию степени выраженности или состояний качественных признаков и диагностически значимых интервалов количественных признаков. Если имеются два класса заболеваний или более, то численное кодирование нужно выполнить так, чтобы сохранить максимальное различие между изображениями разных классов. Это связано с тем, что признаки несут некоторый содержательный смысл, что очень важно, так как при кодировании происходит замена реальных объектов числами (векторами). Примером может служить выборка, составленная из реальных объектов, где в один класс попали все объекты, у которых значение одного из признаков не превосходит 4, а в другой те, у которых значение этого параметра лежит между 5 и 10. При произвольном кодировании можно закодировать объекты таким образом, что это очень важное для распознавания свойство будет утеряно.

Для кодирования признаков I группы область изменения каждого признака разбивается на несколько интервалов (градаций) в зависимости от требуемой степени детализации описания признака и поставленной задачи.

Признаки, входящие во II группу, разделяются на градации в соответствии с изменением их выраженности либо в возрастающем, либо в нисходящем ряду значений симптомов.

Признакам III группы кодовые значения могут быть присвоены произвольно, в соответствии с принятым порядком перечисления показателей. Таким путем осуществляется построение классификационно-квантифицированных шкал признаков, т.е. процедура идентификации и установления физических границ изучаемых параметров. Необходимым условием является монотонность изменения кодов при сохранении максимальных различий между анализируемыми классами (рядами данных).

Находит применение как 4-градационная система, которая соответствует принятой, в клинике для оценки степени выраженности (отсутствие патологического признака или норма, слабая степень, средняя степень и сильная степень проявления признаков), так и

система, включающая до 7 градаций. Последнее позволяет обеспечить также объединение в одном признаке (в виде отдельных градаций), например, различных изменений со стороны сердца (норма — 0, функциональные изменения при отсутствии шума — 2 и при наличии шума — 4, дисфункция митрального клапана — 5, органический порок сердца — 6).

В качестве общего правила можно рассматривать стремление к сокращению числа кодовых чисел за счет внутренних зависимостей или формирование новых признаков. Наиболее целесообразной представляется 7-балльная шкала, что позволяет продемонстрировать дискретность проявления качественных признаков.

Кроме того, учет именно такого количества градаций соответствует возможностям человека по переработке информации, что сформулировано специалистами в области инженерной психологии в виде известного закона «семь плюс минус два» (Miller G., 1956).

Работа с пропущенными и неполными данными

В случае пропусков информации (незамеренные признаки при отсутствии данных об их значениях) применяются различные подходы — усреднение, «обнуление», восстановление или условное кодирование для «обхода» данного значения при статистической обработке.

Использование так называемого «обнуления», т.е. приписывание кодового числа 0, представляется крайне нежелательным, так как это в большинстве случаев совпадает с кодированием нормы. При использовании средних значений число ошибок увеличивается с ростом пропусков информации в такой степени, что приводит к невозможности выбора существенных признаков, в особенности для неоднородных выборок, имеющих «выскакивающие» наблюдения.

Ориентация на средние значения возможна только при выделении очень однородных подгрупп. Наиболее целесообразным следует признать заполнение пропусков предположительно соответствующими им значениями путем, например, их оценивания с помощью регрессии на присутствующие для анализируемого объекта переменные, что детально изложено в монографии

Р. Литтла и Д.В. Рубина (1991). Фактически все подобные подходы основаны на использовании «избыточной» информации, обусловленной наличием связей между исходными признаками.

В случае обработки данных при наличии относительно небольшого количества случайным образом распределенных пропущенных данных (под «небольшим» следует понимать различное их число в зависимости от числа наблюдений, решаемой задачи и метода статистической обработки данных), выделяют фрагмент исходной матрицы данных, не имеющий пропусков

При осуществлении диагностической процедуры наряду с достоверными приходится принимать во внимание и так называемые вероятные симптомы, представляющие собой анамнестические сведения и субъективные оценки врачей. Для более эффективного их использования можно вводить количественную оценку их значимости (по мнению исследователя) путем использования весовых коэффициентов (до 10).

ПРЕДВАРИТЕЛЬНАЯ ОЦЕНКА И ВЫБОР СТАТИСТИЧЕСКИХ МЕТОДОВ

Значительное число статистических методов предполагает нормальный характер распределения вероятностей по Гауссу—Лапласу. Однако именно медицинские данные часто характеризуются отклонением от нормального закона, в связи с чем их оценка по среднему арифметическому и его отклонениям не дает правильной информации. В этом случае может быть использован метод квантильных шкал, который состоит из следующего:

- ранжировка вариационного ряда,
- выделение минимальных и максимальных значений ряда;
- определение значений показателя, составляющих интересующую

исследователя долю (в квантильной шкале) или процент (в центильной шкале) вариационного ряда.

Как правило, прежде чем приступать к решению непосредственно стоящих перед исследователем задач, необходимо осуществить предварительный анализ материала, что предполагает определенную последовательность применяемых методов, начиная с оценки особенностей распределения. Так, например, биномиальное распределение может служить косвенным указанием на возможность гетерогенности исследуемой группы.

Характер распределения необходимо учитывать уже на первом этапе представления данных для статистической обработки, при бимодальном бывает целесообразно ориентироваться не на абсолютные значения переменных, а на степень отклонения их от нормы, так называемые, центрированные переменные. Поэтому следует или осуществить проверку нормальности или сразу выбрать метод, ориентированный на возможность отклонений от этого закона. Это относится уже к сравнению выборок с целью оценки статистической значимости их различий, но не касается частотного анализа (оценки) взаимосвязей признаков.

В связи с этим следует иметь в виду, что распространенный критерий Стьюдента основан на предположении о нормальном распределении вероятностей. Однако когда распределение не соответствует закону Гаусса-Лапласа или это неизвестно исследователю, должны применяться непараметрические критерии, чтобы избежать ошибочных заключений о сходстве или различии сравниваемых групп (выборок). Их выбор зависит также от объема выборки, что является одним из критериев (Гублер Е., 1978).

Новый метод обработки статистических данных бутст-рэп (bootstrap — самообеспечивающийся) не требует информации о виде закона распределения изучаемой случайной величины и в этом смысле может рассматриваться как непараметрический, т.е. работает без опоры на существенную часть априорной информации. Он отличается от традиционных тем, что предполагает многократную обработку различных частей одних и тех же данных, как бы поворот их «разными гранями», и сопоставление полученных таким образом результатов.

Бутст-рэп - процедура может рассматриваться как способ управления выборкой в ходе обработки данных, что является переносом идей активного эксперимента на процедуры вычислений (Эфрон Б., 1988).

Анализ связей

Анализ связей между признаками предполагает:

во-первых, выбор метода с учетом характера анализируемых признаков (количественные, качественные, классификационные);

во-вторых, оценку числового значения по имеющимся данным;

в-третьих, проверку гипотезы, что полученное значение свидетельствует о наличии статистической связи (достоверность корреляционной характеристики с учетом доверительного интервала, покрывающего и вероятные неизвестные значения параметра);

в-четвертых, анализ структуры связей между компонентами многомерного признака.

Под структурой понимается характер связей (непосредственных и опосредованных, тесных и слабых и т.д.) между всеми признаками системы, и ее можно представить в виде графа, где каждая вершина (узел) соответствует одному из признаков, а ребра указывают на определенную связь между признаками

Но коэффициент корреляции имеет четкий смысл как характеристика степени тесноты связи только в случае совместной нормальной распределённости исследуемых признаков. Однако некоррелированность не может служить однозначным указанием на

отсутствие связи между признаками, а лишь указывает на отсутствие линейной зависимости между ними (при этом можно обращаться к использованию индекса корреляции). В

таком случае может идти речь о косвенной связи анализируемых параметров через третьи признаки, которые могут быть учтены при использовании методов многомерного корреляционного анализа. Также и наличие корреляции не является однозначным указанием на их взаимообусловленность (в последнем случае говорят о ложной корреляции).

Для наглядного представления корреляционной зависимости между показателями может быть рассчитана матрица корреляции и по ней построено дерево связей по методу корреляционных плетей (Жилинская М. и Овсенева Т., 1977).

Следует четко различать методы для исследования количественных признаков и ранжированных (порядковая шкала). В последнем случае используются ранговые коэффициенты корреляции Спирмена и Кендалла. Многомерные выборочные распределения, измеренные в номинальных и порядковых шкалах, могут быть проанализированы на предмет имеющихся связей с помощью таблиц сопряженности (contingency tables) представляющих собой прямоугольные таблицы с двух-, трех- или многосторонней классификацией, позволяющие фиксировать число объектов, одновременно имеющих соответствующие градации каждого признака.

Подбор информативных показателей

Поиск наиболее информативной системы показателей может осуществляться при использовании методов классификации многомерных наблюдений. Многочисленная группа методов объединяется общим понятием «анализ данных». Речь идет об агрегировании (сжатии) эмпирических данных с целью приведения их к компактному и обозримому для дальнейшего исследования виду. Существует ряд подходов, различающихся по исходной идее и методам реализации задачи:

- 1) снижение размерности описания объектов путем замены значительного количества исходных показателей небольшим числом интегральных (обобщенных), сохраняющих достаточную информацию об исследуемых объектах (например, главные компоненты);
- 2) факторный анализ, предполагающий, что исходные показатели являются проявлениями небольшого числа объективно существующих, но не поддающихся непосредственному измерению факторов, детерминирующих различия между объектами,
- 3) оценки относительной значимости в тех случаях, когда показатели носят оценочный характер;
- 4) агрегированные показатели или функции от исходных на основе регрессионного, дискриминантного анализа или в виде «произвольных» векторов при многомерном шкалировании;
- 5) функциональное шкалирование (Авен П. и соавт., 1988) путем построения единственного (интегрального) показателя.

Различные методы анализа данных

Многомерное шкалирование — метод построения конфигурации точек в пространстве небольшой размерности исходя из расстояний между ними, искаженных случайными ошибками, или из ранговой информации об этих расстояниях. Этот редко применяемый у нас метод подробно изложен М.Дэйвисоном(1988).

Основной тип данных в многомерном шкалировании — меры близости между двумя объектами (их можно исследовать также с помощью иерархического кластер-анализа). Также как факторный и кластерный анализ, метод используется для описания структуры данных и применим в ряде случаев, когда непригодно большинство методов факторизации. Так, в многомерном шкалировании оценки координат являются непрерывными переменными, тогда как в кластерном анализе — дискретными, а связь между данными о близости и расстоянии не может быть представлена линейной или монотонной функцией.

Факторный анализ также позволяет получить количественное координатное

представление структуры взаимосвязей объектов, однако эти методы опираются на разные предположения о связи между координатами и полученными данными о близости. Необходимо отметить, что выводы, полученные на основе этих двух методов, могут различаться.

Модель факторного анализа предполагает, что структура связей между анализируемыми признаками может быть объяснена тем, что эти переменные зависят от меньшего числа непосредственно не измеряемых («скрытых») факторов. Эта же идея лежит в основе метода главных компонент, также обеспечивающего сжатие информации и одновременно являющегося способом классификации многомерных наблюдений.

Кластер-анализ позволяет обоснованным образом разделять объекты на группы. Обратная процедура агломеративно-иерархического кластерного анализа дает возможность последовательно объединять признаки или объекты на основе вычисления различия или сходства между ними. Таким путем формируются группы, в рамках которых, на основе содержательных соображений, могут быть выделены подгруппы близких (в определенном смысле) объектов. Характерной особенностью является то, что «численная классификация», или таксономия не занимается распределением объектов по известным классам, как методы дискриминантного анализа, а устанавливает новую классификацию.

В множественном регрессионном анализе рассматривается связь между одной переменной, называемой зависимой, и несколькими другими, называемыми независимыми. Эта связь представляется с помощью статистической модели, т.е. уравнения, которое связывает зависимую переменную с независимыми. Функция может быть линейна относительно параметров - линейная модель регрессии, в противном случае — нелинейная. Таким путем могут быть установлены зависимости заболеваемости от характера и уровня химического загрязнения окружающей среды, позволяющие

оценивать вклад различных веществ и их сочетаний, в том числе выбросов конкретных предприятий, в реализацию заболеваний. Однако нужно иметь в виду, что если независимые переменные (аргументы) сильно коррелированы и, соответственно, полученные коэффициенты также сильно коррелированы, то нельзя менять один отдельный фактор, используя соответствующий ему коэффициент.

Дисперсионный анализ позволяет определить влияние разных факторов (условий) на исследуемый признак (явление), что достигается путем разложения совокупной изменчивости (дисперсии, выраженной в сумме квадратов отклонений от общего среднего) на отдельные компоненты, вызванные влиянием различных источников изменчивости. Содержательный анализ исследования различных зависимостей и статистический аппарат их реализации подробно изложены в монографии С.А. Айвазяна и соавт. (1985).

Методы анализа данных нашли широкое распространение, в том числе в исследованиях по изучению угрозы заболевания при наличии факторов риска. Используются и специально разработанные математические модели. Концепция относительного риска рассматривает отношение между пациентами с определенной болезнью и не имеющими ее. Величина относительного риска (r') дает возможность определить, во сколько раз увеличивается вероятность заболеть при его наличии, что может быть оценено с помощью следующей упрощенной формулы (Двойрин В., 1975):

$$r' = \frac{a \cdot b}{d \cdot c},$$

где a — наличие признака в исследуемой группе;

b — отсутствие признака в исследуемой группе;

c — наличие признака в группе сравнения (контрольной);

d — отсутствие признака в группе сравнения (контрольной).

Показатель атрибутивного риска (r) служит для оценки доли заболеваемости, связанной с данным фактором риска:

$$r_a = \frac{Q(r'-V)}{Q(r'-1)+1}$$

где Q - частота признака, маркирующего риск, в популяции;
r'-относительный риск.

Примеры применения выше приведенных формул для расчета рисков можно найти в монографии «Эпидемиология неинфекционных заболеваний» (Вихерт А. и соавт., 1990). В современных зарубежных руководствах по эпидемиологии (Norell S. et al., 1995) рассматривается также возможность оценки относительного риска при определенных условиях с использованием показателя отношения вероятностей заболевания (odds ratio).

Выявление факторов, способствующих возникновению (проявлению) заболевания, т.е. факторов риска, может осуществляться различными способами, например, путем оценки информативности с последующим ранжированием признаков, что, однако, не указывает на совокупное действие отобранных параметров в отличие от применения регрессионного, факторного анализа, методов теории распознавания образов, которые дают возможность получать «симптомокомплексы» риск-факторов. Кроме того, более сложные методы позволяют анализировать и не прямые связи между факторами риска и заболеваниями.

Дискриминантный анализ обобщенное название методов классификации объектов, представленных многомерными наблюдениями. Осуществляется в медицине в целях дифференциальной диагностики, прогнозирования осложнений или течения заболевания. В тех случаях, когда это осуществляется на основе имеющихся в распоряжении исследователя выборок, чаще используются методы теории распознавания образов на основе стохастических (вероятностных) процедур или детерминистской логики.

Фактически основой первого класса подходов является процедура Байеса, которая позволяет выбрать одну из нескольких гипотез (диагностических, прогностических) на основе вычисления вероятностей болезней (исходов) по вероятностям отмеченных у больных признаков. Однако классическая байесовская формула предполагает независимость признаков, что необходимо учитывать при ее использовании, осуществляя предварительный отбор симптомов, или использовать упрощенную формулу для случая зависимых признаков. Последовательный анализ Вальда и неоднородная последовательная процедура (Гублер Е., 1978) позволяют при диагностике или прогнозировании получать решение об отнесении больного к тому или иному классу непосредственно по достижении определенного порога, предварительно установленного на обучающей выборке с известными результатами. Таким путем достигается возможность различия проверяемых гипотез с заданной точностью при значительно меньшем числе наблюдений.

Для выявления оптимального сочетания признаков, имеющих диагностическое (прогностическое) значение, могут применяться различные методы теории распознавания образов, основанные на вычислении расстояний между признаками, характеризующими больных сравниваемых классов. По при этом следует учитывать соотношения объекты — симптомы, считая идеальным 10:1, оптимальным 5:1 (эти варианты теоретически обоснованы, но на практике используются сравнительно редко) и минимальным 2:1, имея в виду, что это отражается на надежности получаемых результатов.

Целесообразно обратить внимание на визуальную форму представления результатов. Это могут быть таблицы, графики, гистограммы, геометрическая интерпретация классификации многомерных наблюдений. Имеются методы, априори предполагающие такой вид выходных форм, например, главные компоненты, кластер-анализ. В других случаях можно ставить такую задачу и реализовывать ее с помощью специально написанного программного обеспечения.

ЛИТЕРАТУРА

1 Авен П О, Ослон А А, Мучник И Б Функциональное шкалирование -М, 1988 -182с

- 2 Айвазян С А Енюков И О Мешалкин ЛД Прикладная статистика Основы моделированной первичной обработки данных Справочное издание Финансы и статистика - М , 1983 - 471 с
- 3 Айвазян С А, Енюков И О, Мешалкин ЛД Прикладная статистика Исследование зависимостей Справочное издание - М , Финансы и статистика 1985 -487с
- 4 Амосов Н.М, Зайцев НГ, Мельников В Г и др Медицинская информационная система -Киев, 1975
- 5 Вихерт АМ, Жданов В С, Чаклин А В и др Эпидемиология неинфекционных заболеваний - М , 1990 - 272 с
- 6 Гублер ЕВ Вычислительные методы анализа и распознавания патологических процессов - Л , 1978 - 294 с
- 7 Двойрин В В Методы эпидемиологических исследований при злокачественных опухолях -М, 1975 - 100с
- 8 Дэвисон М Многомерное шкалирование методы наглядного представления данных - М Финансы и статистика, 1988 - 254 с
- 9 Жилинская М.В. Овсенева Т.Д Метод корреляционных плеяд в изучении структуры связей показателей эритропитарной системы при пиелонефрите / Медицинская и биологическая кибернетика - М , 1977, Т 2 -С 145- 150
- 10 Кобринский Б А Принципы математико-статистического анализа данных медико-биологических исследований // Росс вест перинат и пед -1996 -№4 -60-64 с
- 11 Литгл РДжА, Рубин ДБ Статистический анализ данных с пропусками Финансы и статистика - М , 1991 - 336 с
- 12 ЛюбищевАА Дисперсионный анализ в биологии -М,1986 -200с
- 13 Эфрон К Нетрадиционные методы многомерного статистического анализа Финансы и статистика - М , 1988 - 263 с
- 14 Miller G A The magical number seven, plus or minus two, some limits on our capacity, for processing information//Psichol Rev -1956 Vol 63 -P 81-87
- 15 Noell S.E 5 E Workbook of Epidemiology - Oxford Unit Press, 1995 -317p