ШКАЛЫ ИЗМЕРЕНИЙ

Слайд 2

Основное понятие в системах обработки информации – данные, которые подразделяются на *качественные* (*нечисловые*) и *количественные* (*числовые*). Примерами могут служить: пол, социальное положение и температура, вес, рост человека. Данные получаются в результате измерений признаков индивидуумов или подопытных образцов из исследуемой совокупности. Измерение – присвоение символов подопытным образцам в соответствии с некоторым правилом. Для измерения важны аксиомы сравнения величин:

Слайд 3

Аксиомы тождества:

1. Либо $A = B$, либо $A \neq B$.
2. Если $A = B$, то $B = A$.
3. Если $A = B$ и $B = C$, то $A = C$.

Аксиомы рангового порядка:

4. Если $A > B$, то $B < A$.
5. Если $A > B$ и $B > C$, то $A > C$.

Аксиомы аддитивности:

6. Если $A = P$ и $B > 0$, то $A + B > P$.
7. $A + B = B + A$.
8. Если $A = P$ и $B = Q$, то $A + B = P + Q$.
9. $(A + B) + C = A + (B + C)$.

Слайд 4

На этой аксиоматической основе строится теория шкал измерений. Для качественных измерений различают типы шкал: *наименований* (*номинальная*), *порядковая* (*ординальная*) и *гиперпорядка*. В количественных измерениях выделяются четыре шкалы: *интервальная, отношений, разностей*, *абсолютная*. Тип шкалы определяется группой допустимых преобразований $\mathbf{\Phi} = \{\varphi(x)\}$ шкалы, переводящих одну систему измерений, являющейся гомоморфным образом эмпирической системы, в другую, также являющейся гомоморфным образом этой же эмпирической системы.

К числу преобразований, характеризующих основные типы шкал, относятся: взаимно-однозначные, монотонные, тождественные, подобия, сдвига, линейные преобразования. Чем меньше множество систем, в которые гомоморфно отображается рассматриваемая эмпирическая система с отношением, тем сильнее шкала. Тип шкалы также определяет возможности применения к измерениям операций сравнения, арифметических действий.

Слайд 5

Качественные измерения

Шкала наименований. Логическая основа шкалы содержится в первых трех аксиомах. Объекты можно опознавать и различать на основе этих аксиом идентификации. Построение шкалы наименований означает, что измерения разбиваются на классы по определенному признаку. Измерениям, попавшим в один класс, соответствует одно обозначение. Аксиомы тождества можно использовать для введения бесконечного набора различимых названий классов. Например, мужской и женский пол можно обозначить *М* и *Ж*, или 1 и 2 (или 2 и 1), или знаками небесных светил ♂ и ♀. Разбиение на классы необходимо осуществлять так, чтобы различия внутри класса были малы, а между классами – велики. При этом классы должны иметь неупорядоченный характер и "не перекрывать" друг друга. Отсюда следует, что процедура *неупорядоченной классификации* может рассматриваться как измерение по шкале наименований.

Группа допустимых преобразований в этой шкале состоит из *всех* взаимно однозначных преобразований. Арифметические операции не имеют смысла. Нет смысла арифметического среднего, медианы. В качестве параметра сдвига (среднего) может служить мода распределения, которая не зависит от однозначных преобразований измерений. Например, гипертоников больше, чем гипотоников. Мода описывает параметр сдвига для гипертоников и гипотоников вне зависимости от обозначений: ГП и ГИП, 1 и 2. Хотя измерения качественны, но можно сосчитать число объектов каждого класса и определить частоты.

Методы анализа: пригодны только методы категориального анализа – хи-квадрат критерий для полиномиального распределения, хи-квадрат для проверки гипотезы о связанности двух и более номинальных переменных, выводы относительно биномиального распределения, операции над структурными функциями от дихотомических переменных.

Слайд 6

Порядковая шкала. В этой шкале качественные измерения разбиваются не только на классы, но и упорядочиваются классы. Каждому классу соответствует собственный символ и порядок символов соответствует порядку класса по правилу «больше, чем», «меньше, чем», «более предпочтителен, чем», «сильнее». Любое множество A называется упорядоченным, если для любых двух его элементов A и B установлено, что, либо A предшествует B, либо B предшествует A (аксиома 4). Иногда не удается установить строгое предшествование для всех элементов множества, однако можно выполнить «групповое» упорядочение, когда упорядочиваются подмножества равноценных элементов (аксиома 5).

Далее можно поставить задачу сравнения и упорядочения этих множеств. Аксиомы 4 и 5 определяют сравнимость и транзитивность элементов порядковой шкалы по некоторому общему признаку. Например, можно упорядочить индивидуумы по цвету волос (x – светлый, y – серый, z – черный); социальное положение по уровню доходов – низкий, средний, высокий (можно эти классы обозначить: 0, 1, 10 или 10, 1, 0, или а, б, с).

В порядковых шкалах нельзя определить *меру доминирования*, т.е. измерить, *насколько* один объект лучше, важнее другого. Не всегда ясен принцип сравнения и возникают случаи нетранзитивности. Например, команда A выигрывает у команды B, команда B выигрывает у команды C, но команда C выигрывает у A.

Различают также в этой шкале данные, образующие *слабый порядок.* Для элементов шкалы слабого порядка изменяются аксиомы 4 и 5 следующим образом:

4'. Либо $A \geq B$, либо $A \leq B$.

5'. Если $A \geq B$ и $B \geq C$, то $A \geq C$.

Если ни A ≥ B, ни A ≤ B, то говорят, что *A* и *B несравнимы.* Если одновременно A ≥ B и A ≤ B, то получаем рефлексивное отношение A = B. Если некоторые элементы измеряемого множества несравнимы по упорядочивающему отношению, а остальное подмножество элементов допускает сравнение, то имеем так называемый *частичный порядок*.

Преобразования в шкале порядка – *монотонные функции* $\varphi(x)$. Использование порядковых шкал позволяет производить все операции номинальной шкалы и преобразования полученных оценок (например, от экспертов), которые отвечают всем монотонно возрастающим функциям. Так, например, положительные оценки могут быть заменены их квадратами, или логарифмами, или любой другой монотонно возрастающей функцией.

Слайд 7

Арифметические операции не имеют смысла. Нельзя за среднее взять среднеарифметическое. Характеристикой сдвига могут служить медиана или с некоторой потерей информации, мода распределения.

Методы анализа: пригодны *непараметрические методы*, не зависящие от параметрического семейства распределений. Эти методы используют знаки, ранги абсолютных значений и знаки их разностей, число инверсий. Применяются следующие процедуры: проверки гипотезы равенства медианы заданному значению, равенства двух медиан, методы дисперсионного анализа Крускала-Уоллиса и Фридмана, методы ранговой корреляции Кендалла и Спирмена.

Слайд 8

***Примеры* 1.1.** Порядковые шкалы: в медицине – шкала стадий гипертонической болезни по Мясникову, шкала степени выраженности коронарной недостаточности по Фогельсону; в минералогии – шкала твердости минералов Мооса (тальк – 1, гипс – 2, кальций – 3, флюорит – 4, апатит – 5, ортоклаз – 6, кварц – 7, топаз – 8, корунд – 9, алмаз – 10), в географии – бофортова шкала ветров («штиль», «слабый ветер», «умеренный ветер»). ■

Шкала гиперпорядка. Шкала гиперпорядка отличается от порядковой шкалы тем, что в ней допустимы гипермонотонные преобразования. Преобразования $\varphi(x)$ являются гипермонотонными, если для любых x, y, u, v

$$\varphi(x)-\varphi(y)<\varphi(u)-\varphi(v)$$

только, когда x, y, u, v принадлежат области определения $\varphi(x)$ и $x - y < u - v$. Измерения в шкалах гиперпорядка сохраняют упорядочение разностей численных оценок. Любые линейные преобразования являются гипермонотонными: но обратное неверно.

Слайд 9

Количественные измерения

Интервальная шкала. В интервальной шкале численные значения числовой системы измерений определяются с точностью до линейных преобразований

$$\varphi(x)=\alpha x+\beta, \quad \alpha>0.$$

Начало отсчета и единица измерения в этой шкале не фиксированы, и при переходе от одной системы измерений к другой они должны быть заданы. Шкала позволяет не только указать, что один объект больше другого в принятой единице измерения, но и утверждать, насколько велики расстояния между двумя точками в шкале. Интервальная шкала – это порядковая шкала плюс известные расстояния между любыми двумя числами на шкале. В общем случае: нуль и единица измерения шкалы выбираются произвольно.

***Пример* 1.2.** Измерение температуры по Цельсию, Фаренгейту. Если x – температура по Фаренгейту, y – по Цельсию, то

$$y = (5/9)(x - 32) = 0{,}55x - 17{,}8.$$

Здесь можно не только утверждать, что температура 50 градусов выше, чем температура 20 градусов, но и то, что увеличение температуры с 10 до 50 градусов в пять раз больше увеличения температуры с 40 до 50 градусов. ■

Слайд 10

В этой шкале отношение оценок не сохраняется, т.е. отношение одного измерения к другому изменяется вместе с типом используемой системы измерения. Например, температурам 100° и 0°С в шкале по Цельсию соответствуют температурам 212° и 32°F по Фаренгейту. Однако в шкале интервалов сохраняются отношения разностей численных оценок. Действительно, пусть x_1, x_2, x_3, x_4 – измерения признака в некоторой числовой системе, а $\varphi(x_1)$, $\varphi(x_2)$, $\varphi(x_3)$, $\varphi(x_4)$ – соответствующие им измерения этого же признака в другой числовой системе. Тогда

$$\frac{\varphi(x_1) - \varphi(x_2)}{\varphi(x_3) - \varphi(x_4)} = \frac{\alpha x_1 - \alpha x_2}{\alpha x_3 - \alpha x_4} = \frac{x_1 - x_2}{x_3 - x_4}.$$

Шкалы интервалов не обладают важными свойствами *аддитивности*, которые определяются аксиомами 6 – 9. Это означает, что для этих шкал *не адекватны* (неосмысленны) числовые утверждения, но осмысленны числовые операции. Например, в интервальной шкале утверждение 2 + 3 = 5 бессмысленно, так как

$$\alpha 2 + \beta + \alpha 3 + \beta \neq \alpha 5 + \beta \quad \text{при } \beta \neq 0,$$

но утверждение 2 + 3 = 4 + 1 осмысленно.

Слайд 11

В интервальной шкале применимы оценки среднеарифметического, дисперсии, высших и смешанных моментов, процедуры сравнения средних арифметических для двух совокупностей объектов. Действительно, если x_i, $i = 1, \ldots, n$ и y_i, $i = 1, \ldots, m$, то высказывание

$$(1/n)\sum_{i=1}^{n} x_i > (1/m)\sum_{i=1}^{m} y_i$$

эквивалентно утверждению

$$\frac{1}{n}\sum_{i=1}^{n}(\alpha x_i + \beta) = \frac{\alpha}{n}\sum_{i=1}^{n} x_i + \beta > \frac{1}{m}\sum_{i=1}^{m}(\alpha y_i + \beta) = \frac{\alpha}{m}\sum_{i=1}^{n} y_i + \beta.$$

Слайд 12

Однако отношение среднеарифметических величин не имеет смысла. Корректно лишь использование отношения разностей средних арифметических. Отсюда следует, что не имеет смысла коэффициент вариации – отношение среднеквадратического отклонения к математическому ожиданию, так как обе эти величины зависят от начала отсчета.

Слайд 13

Шкала отношений. Шкалой отношений называется шкала, в которой численные значения числовой системы определяются с точностью до преобразований подобия

$\varphi(x) = \alpha x,\ \alpha > 0$.

Здесь, в отличие от интервальной шкалы, фиксировано начало отсчета, но не единица измерения.

Свойство шкалы: при выборе любой системы измерений, отношение измерений одинаково при переходе от одной системы к другой. Пусть измерения x_1 и x_2 в первой системе, а $y_1 = \alpha x_1$, $y_2 = \alpha x_2$ – в другой. Тогда $x_1 / x_2 = y_1 / y_2$, то есть отношение любых двух точек шкалы не зависит от единицы измерения.

Эта шкала имеет все свойства шкал количественных измерений и для нее справедливы свойства аддитивности, определяемые аксиомами 6 – 9. Шкала отношений обычная в технических и физических системах, но редко встречается в обществоведении и гуманитарных дисциплинах.

Слайд 14

Шкала разностей. В этой шкале единица измерения фиксирована, а начало отсчета – нет, т.е. при переходе от одной числовой системы к другой изменяется лишь начало отсчета. Преобразование: $\varphi(x) = x + \beta$.

Примеры. Солнечные календари: Юлианский, Григорианский, Древнерусский, Великой Французской революции, Всемирный. ■

Слайд 15

Абсолютная шкала. Известны начало отсчета и единица измерения. Численные значения измерений равны тождественно значениям числовой оси. Например, количество студентов в аудитории.

Преобразование – тождественное, $\varphi(x) = x$. Для этой шкалы равно, как и для любых шкал количественных измерений, имеют смысл все арифметические операции. В качестве оценки параметра сдвига могут использоваться среднее, медиана, мода распределения.

Слайд 16

Преобразования шкал. Рассмотренные типы шкал определяются допустимыми преобразованиями. В табл. 1 приведены типы шкал с определяющими их классами допустимых преобразований.

Типы качественных и количественных шкал Таблица 1.

Измерения	Шкала	Допустимые преобразования
Качественные	Наименований Порядковая Гиперпорядка	$\varphi(x)$ – взаимно однозначные $\varphi(x)$ – монотонные $\varphi(x)$ – монотонные, сохраняющие порядок первых разностей
Количественные	Интервальная Отношения Разностей Абсолютная	$\varphi(x) = \alpha x + \beta,\ \alpha > 0$ $\varphi(x) = \alpha x,\ \alpha > 0$ $\varphi(x) = x + \beta$ $\varphi(x) = x$

В табл. 1. шкалы расположены по мере увеличения их силы. Из этих шкал абсолютная является самой сильной, а шкала наименований – самой слабой. Однако использовать только сильные шкалы невозможно, так как процедур и приборов для измерения многих характеристик в сильных шкалах не существует. У слабых шкал имеются и некоторые преимущества. Они более помехоустойчивы, если помеха находится в рамках допустимых преобразований. При этом группа допустимых преобразований тем шире, чем слабее шкала.

Задача преобразования шкал не является простой. Рассмотрим пример измерения мощностей системы на входе и выходе. В большинстве случаев для измерения используется отношение мощностей на выходе и на входе системы в логарифмической шкале в децибелах

$N = 10\, ln(P_{\text{вых}}/P_{\text{вх}})$,

где $P_{\text{вых}}$ и $P_{\text{вх}}$ измеряются в шкале отношений и взятие логарифма допустимо. В результате измерения в децибелах получается не шкала отношений, а шкала интервалов, следовательно, при использовании шкалы в децибелах нужно соблюдать законность операций интервальной шкалы. Например, бессмысленно говорить о мощности в децибелах в какой-либо точке системы без указания мощности в некоторой опорной точке системы.

Методы вычисления, справедливые в слабой шкале, применимы, с некоторой потерей информации, и для наблюдений в сильных шкалах. Обратное утверждение не имеет смысла, так как данные слабой шкалы невозможно обработать методами сильных шкал.

Квазиколичественные измерения

Качественные данные (например, многозначные номинальные; порядковые данные; данные, построенные на субъективных критериях) в практических приложениях удобно заменять квазиколичественными переменными, называемыми *квазиквантитативными* [1]. При такой замене порядковых данных происходит некоторая потеря информации, но зато модели на квазиквантитативных данных ничем не отличаются от моделей для количественных данных. Каждый качественный признак можно представить в виде квазиквантитативного признака.

Пусть имеется одно измерение x качественного признака X с k значениями классов. Предположим, что наблюдения классов подвергнуты некоторому упорядочению, достигнутому, например, путем соглашения. Тогда измерение x

[1] Квантитативное – количественное (от лат. quantitas – количество).

можно представить как квазиквантитативную переменную, определяемую как k-мерный вектор-строку (подвектор) с компонентами, принимающими только значения 0 и 1, причем одна компонента вектора, соответствующая наблюдаемому классу, равна 1, а остальные компоненты равны 0.

Если значения признака измерены у n объектов (n-подвек-торов), то получим матрицу квазиквантитативной переменной $\mathbf{X}$ размерности $n \times k$ с нулевыми и единичными элементами. В этой матрице каждый объект отображается своей строкой как точка k-мерного пространства, а признак отображается не одним столбцом, а k столбцами:

$$\mathbf{X} = [(x_{11}, \ldots, x_{1k}), (x_{21}, \ldots, x_{2k}), \ldots, (x_{n1}, \ldots, x_{nk})]',$$

где

$$x_{ji} = \begin{cases} 1, & \text{если событие } s_i \ (i = 1, \ldots, k) \text{ наступило в } j-\text{м подвекторе;} \\ 0 - \text{в противном случае.} \end{cases}$$

***Пример* 1.** Пусть дан вектор-столбец качественных измерений в порядковой шкале признака X: $X = [1\ 2\ 1\ 3]'$, $k = 3$, $n = 4$. Преобразовав эти данные в квазиквантитативную переменную, получим матрицу размерности (4×3):

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Так как для каждого подвектора матрицы $\mathbf{X}$ справедливо соотношение $\Sigma^{k}_{s=1} x_{js} = 1$, то представление качественных данных с помощью квазиквантитативных переменных избыточно. Наблюдение *дихотомической* (*булевской*, *бинарной*) переменной, состоящей всего из двух элементов «0 или 1», соподчиненных постулату классического принципа исключенного третьего: «*А* или *не-А*», является частным случаем квазиквантитативной переменной при $k = 2$ для всех j. ■

***Пример* 2.**

1. Дихотомической переменной x_i может служить состояние i-го элемента системы, определяющее, находится ли элемент в работоспособном или не в работоспособном (отказном) состоянии:

$$x_j = \begin{cases} 1, & \text{если } i-\text{й элемент работоспособен,} \\ 0, & \text{если } i-\text{й элемент отказал.} \end{cases}$$

Подобная дихотомия применима и к системе в целом, работоспособность которой можно описать дихотомической функцией $\varphi(\mathbf{x})$:

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{если система работоспособна,} \\ 0, & \text{если система отказала.} \end{cases}$$

где $\mathbf{x} = (x_1, \ldots, x_n)$ вектор состояний элементов.

2. Контроль качества продукции по альтернативному признаку означает, что единица продукции относится к одной из двух категорий – «годных» или «дефектных».

3. Особенно большое применение дихотомические переменные имеют в медико-биологических и социологических исследованиях, в которых большинство переменных, интересующих специалистов, не может быть измерено в

количественных шкалах. При этом дихотомические данные зачастую являются более адекватными, чем результаты измерений по методикам, использующим большее число градаций. Некоторые психологические тесты используют только дихотомические данные. На дихотомические данные опираются и методы парных сравнений.

Любой номинальный признак можно преобразовать в дихотомическую переменную, но не наоборот. Это связано, в частности, с тем, что при дихотомическом делении один из классов является дополнительным и определяется всегда только отрицательно, тогда как при номинальном делении оба класса определяются положительно, заменить же отрицательное определение положительным не всегда возможно. Например, множество людей можно делить на *мужчин и не-мужчин* (по признаку быть мужчиной), но деление этого множества на *класс мужчин и класс женщин* по признаку пола не является дихотомическим делением – основания деления здесь разные, свойство «быть мужчиной» не противоречит свойству «быть женщиной».

Взаимно однозначные преобразования для квазиквантитативных переменных совпадают с множеством линейных преобразований. Для квазиквантитативных переменных x_1 и x_2 справедливы преобразования [16]:

$$y_1 = \alpha x_1 + \beta \text{ и } y_2 = \alpha x_2 + \beta ,$$

которые однозначно разрешимы относительно α и β, причем из $\alpha \neq 0$ следует $y_1 \neq y_2$.

Квазиквантитативный признак X качественных измерений с двумя классами задается двумя булевскими столбцами, соответствующими каждому из двух значений классов, что дает возможность линейного представления любого допустимого преобразования. В силу избыточности квазиквантитативный признак с двумя столбцами можно задать единственным столбцом, в котором единицы характеризуют одно, а нули другое значение признака. Такое представление квазиквантитативной переменной с k классами матрицей с $(k - 1)$ столбцами носит название *индикаторной переменной* (*IND*) и широко используется в пакетах прикладных программ при преобразовании качественных переменных в количественные.

Квазиквантитативное представление возможно и для признаков, измеренных в других шкалах. Такое представление может служить в качестве исходных данных для применения линейных методов анализа, так как линейные преобразования квазиквантитативных переменных соответствуют любым, сколь угодно сложным нелинейным преобразованиям исходных переменных.