

Виды данных

План лекции

- Критерии измерений
- Виды данных
- Типы шкал

Измерения и шкалы

- Измерение (measurement) означает присвоение чисел характеристикам изучаемых объектов, явлений согласно некоторому правилу
- Шкала (scale) есть правило (или алгоритм), в соответствии с которым изучаемым объектам, явлениям присваиваются числа



Данные (data)

- результаты наблюдений, испытаний, накапливаемые с целью последующего изучения и анализа.

Респондент	Возраст	Пол	Образование	Семейное положен.
1	29	0 12	2	
2	23	1 14	1	
3	37	1 16	2	
4	46	0 10	4	
5	34	1 14	1	

Переменная

- Переменная - признак (исследуемого объекта), который может принимать разные значения для различных случаев или для различных моментов времени в рамках одного случая.
- Признак - характеристика предмета или явления, по которым его отличают от других предметов или явлений.
- Признак - свойство или характеристика изучаемого явления, выраженные в совокупности переменных, которые могут быть подвергнуты наблюдению и измерению.

Типы признаков:

- Качественные, категориальные:
 - номинальные
 - дихотомические
 - порядковые, ординальные, ранжируемые
- Количественные, интервальные
 - дискретные
 - непрерывные

Номинальная шкала

(nominal scale)

[лат. nominalis именной]

состоит из названий, имен или категорий для сортировки или классификации объектов, явлений по некоторому признаку.

Результаты измерений, полученные при помощи номинальной шкалы, не могут быть упорядочены и с ними не могут производиться арифметические операции.

Приведите примеры

Качественные, категориальные *номинальные*

- ✓ не поддающиеся непосредственному измерению
- ✓ не могут быть упорядочены
- ✓ с ними не могут производиться арифметические операции

диагноз, профессия, семейное положение

Качественные, категориальные *дихотомические (binary)*

Могут быть отнесены только к двум
противоположным категориям,
принимаящие одно из двух значений

Приведите примеры

Дихотомическая шкала (dichotomous scale)

- ✓ Номинальная шкала, которая состоит из двух категорий.

да/нет, жив/умер, курит/не курит

Качественные, категориальные *порядковые (ordinal)*

Можно расположить в естественном порядке, но отсутствует количественная мера расстояния между величинами

Приведите примеры

Порядковая шкала (ordinal scale)

- ✓ числа присваиваются объектам, чтобы обозначить относительные позиции объектов, но не величину различий между ними.

Степень тяжести

Стадия болезни

Самооценка состояния здоровья

Интервальная шкала (interval scale)

позволяет находить разницу между двумя величинами. Обладает всеми свойствами номинальной и порядковой, но она позволяет указать количественное значение измеряемого признака. Недостатком служит отсутствие абсолютного нуля в качестве точки отсчета.

Приведите примеры

Количественные или интервальные

- ✓ Признаки, количественная мера которых четко определена

Т, САД, ДАД, рост, вес, уровень
холестерина, количество беременностей,
дни нетрудоспособности

Какие из них непрерывные, какие – дискретные?

Количественные *непрерывные*

- ✓ Принимают любое значение на непрерывной шкале

Масса тела, температура, биохимические
показатели крови

Количественные *дискретные*

- ✓ Принимают лишь определенные значения из диапазона измерения, обычно целые

Число детей в семье, число заболеваний у одного больного, число выкуриваемых сигарет

Относительная шкала (ratio scale)

- ✓ обладает абсолютным нулем в качестве точки отсчета, что позволяет ей иметь все свойства интервальной шкалы. Для данных этой шкалы осмысленными являются все операции, включая вычитание и дроби.

время на выполнение теста по математике

Шкалы – подведем итог

- Номинальная
 - Содержит только категории, данные не могут упорядочиваться
- Дихотомическая
 - Разновидность номинальной. Содержит всего две категории
- Порядковая
 - Содержит категории, которые могут упорядочиваться, разности не имеют смысла
- Интервальная
 - Разности между значениями могут быть вычислены, но отсутствует точка отсчета
- Относительная
 - Имеется точка отсчета, возможны отношения между значениями

Результаты забега на скачках:

- Дихотомический признак. Эта лошадь пришла первой?
0 – нет, 1- да
- Порядковый. Какой эта лошадь пришла на финиш?
1 – первой, 2 – второй, 3 – третьей и т.д.
- Количественный признак. Какой результат у этой лошади?
60 сек., и т.д.

Производные (вторичные) данные

- **Проценты.** *Состояние больного улучшилось на 24 % после лечения, т.е имеет значение степень улучшения, а не абсолютные данные*
- **Пропорции или отношения.** *Индекс массы тела*
- **Интенсивность.** *Относительная частота заболеваний, где кол-во заболеваний делят на общее число лет, в течение которых вели наблюдения за пациентами*
- **Метки, оценки** *используют, когда невозможно измерить количество. Н-р, опросники качества жизни*

Цензурированные данные

- *Если мы используем прибор, который может обнаружить значения только выше некоторого предельного уровня, тогда любая величина ниже этого уровня не будет обнаружена*
- *Когда некоторые больные из основной группы отстраняются от испытания до окончания исследований*

РЕДУКЦИЯ ДАННЫХ

- уменьшение числа категорий в наборе данных для упрощения анализа.
- суммирование данных посредством классификационных схем и арифметических действий.
- суммирование набора или наборов данных в форме индекса, например, ожидаемой продолжительности жизни или валового внутреннего продукта.

Редукция данных:

Возраст:

Годы (16 лет) – количественный признак

Десятилетия (10-20 лет) – интервальный

Периоды (юношество) – ординальный

Молодой, подросток - номинальный

Какие есть вопросы?