**Типы данных**

Существуют две разновидности случайных переменных,[[1]](#footnote-1) значения которых образуют наборы данных: категорийные и числовые (рис. 1).



Рис. 1. Разновидности данных

**Категорийные случайные величины** возникают в результате ответов на заданные вопросы, скажем, *да* или *нет*. Например, ответить на вопрос «Владеете ли Вы в настоящее время какими-либо акциями или облигациями?» можно лишь положительно или отрицательно. Другим примером подобных данных являются ответы на вопрос о качестве услуг, предоставляемых компанией: «Хотели бы Вы приобрести еще какие-нибудь товары в нашей компании в течение следующих 12 месяцев?». Категорийные переменные могут иметь не только два возможных значения. Например, существуют несколько вариантов ответа на вопрос: «В какой день недели вы предпочитаете обедать в ресторане?».

**Числовые случайные величины** являются ответами на вопросы о каком-либо измерении, например, о росте опрашиваемого. Кроме того, в ответ на вопросы «Сколько денег (в долларах) Вы планируете потратить на приобретение мобильных устройств на протяжении следующих 12 месяцев?» или «На сколько рассылок Вы подписаны?» опрашиваемый также должен указать конкретное число. Существуют две разновидности числовых переменных: дискретные и непрерывные.

**Непрерывная случайная величина** возникает как ответ на вопрос, требующий измерения. Типичным примером такой величины является рост, который может изменяться в определенном интервале и измеряться с заданной точностью. Например, ваш рост может равняться 170 см; 170,3 см; 170,32 см в зависимости от точности проведенных измерений. Теоретически не существует двух людей, имеющих одинаковый рост, поскольку, чем точнее проводятся измерения, тем выше вероятность обнаружить различие между полученными величинами. Однако большинство измерительных приборов не настолько совершенны, чтобы выявлять небольшие различия между измеренными величинами. Поэтому в большинстве случаев результаты эксперимента или опроса содержат взаимосвязанные наблюдения, даже если случайная величина на самом деле является непрерывной.

**Шкалы измерений**

Данные можно классифицировать по шкалам, или уровням измерений. Существуют четыре общепризнанных шкалы измерений: номинальная, порядковая, интервальная и шкала отношений.

**Номинальная и порядковая шкалы.** Данные, представляющие собой значения категорийных переменных, измеряются либо по номинальной, либо по порядковой шкале. Номинальная шкала (рис. 2) классифицирует данные по разным неупорядоченным категориям. Например, ответ на вопрос «Планируете ли Вы приобретать мобильное устройство в течение следующих 12 месяцев?» является номинальной переменной. Аналогично номинальными переменными являются ответы на вопросы о любимых напитках, а также о политической или половой принадлежности. Номинальное шкалирование является слабейшей формой измерения, поскольку исследователи не дифференцируют результаты, принадлежащие одной и той же категории, и не устанавливают отношение порядка между категориями.



Рис. 2. Примеры номинальных шкал

**Порядковая шкала** классифицирует данные по разным упорядоченным категориям. Например, ответ на вопрос «Как Вы оцениваете качество обслуживания Вашего последнего заказа?» представляет собой порядковую переменную, поскольку ее значения ранжируются по степени удовлетворенности клиентов: намного лучше ожидаемого, лучше ожидаемого, соответствует ожиданиям, хуже ожидаемого, намного хуже ожидаемого (рис. 3).



Рис. 3. Примеры порядковых шкал

Порядковая шкала представляет собой более точную форму измерений, поскольку между ответами, отнесенными к разным категориям, устанавливается отношение порядка. Несмотря на это, порядковое шкалирование является разновидностью относительно менее точных измерений, поскольку данные, относящиеся к одной и той же категории по-прежнему не дифференцируются. При порядковых измерениях у исследователей нет разумных инструментов, позволяющих дать количественную оценку ответов. Известно лишь, какая категория *больше*, *лучше* или *предпочтительнее*, но неизвестно насколько.

**Интервальные шкалы и шкалы отношений.** *Интервальная шкала* (рис. 4) представляет собой порядковую шкалу, в которой разности между измерениями выражаются ненулевым числом. Например, температура воздуха, равная 27°С, на 2°С теплее, чем 25°С. Кроме того, разность между температурами, равными 24°С и 26°С, также равна 2°С. Следовательно, указанные разности сохраняют смысл для любых измерений.



Рис. 4. Примеры интервальных шкал и шкал отношений

*Шкала отношений* — это упорядоченная шкала, в которой разности между измерениями (высоты, веса, возраста или зарплаты) могут равняться нулю. Например, сумма денег (в долларах США), которую клиент планирует потратить на приобретение стереофонического оборудования на протяжении следующих 12 месяцев, представляет собой переменную, измеренную по шкале отношений. Кроме того, шкала отношений может содержать рост человека, равный 176 см, который вдвое превышает рост другого человека, равный 88 см. Температура представляет собой более сложный случай: шкалы Фаренгейта и Цельсия являются интервальными, но их нельзя назвать шкалами отношений, поскольку нулевая температура — это условная величина, а не реальная. Нельзя сказать, что температура воздуха, равная 76°F, вдвое теплее, чем температура, равная 38°F. Однако, в отличие от шкал Фаренгейта и Цельсия, шкала Кельвина является шкалой отношений, поскольку включает в себя не условный, а абсолютный ноль.

Значения числовых переменных, как правило, измеряются либо по интервальной шкале, либо по шкале отношений. Эти шкалы образуют высший уровень измерения. Они точнее, чем порядковая шкала, поскольку позволяют определить, не только, какая из наблюдаемых величин больше другой, но и насколько.

**Контрольные задания**

1. Какие случайные величины приведены ниже — категорийные или числовые? Если переменная является числовой, определите ее разновидность (дискретная или непрерывная). Определите уровень измерения.
	1. Количество телефонов в офисе.
	2. Наиболее распространенный тип телефона.
	3. Количество междугородных разговоров за месяц.
	4. Продолжительность (в минутах) наиболее долгого междугородного телефонного разговора за последний месяц.
	5. Наиболее распространенный цвет телефона.
	6. Ежемесячная оплата (в долларах и центах) за междугородные телефонные разговоры.
2. Предположим, что от студентов, посещавших книжный магазин в студенческом городке на протяжении первой недели занятий, получена следующая информация:
	1. Количество денег, потраченных на книги.
	2. Количество приобретенных книг.
	3. Количество времени, проведенного в магазине.
	4. Академическая специализация студента.
	5. Пол.
	6. Владение персональным компьютером.
	7. Количество курсов, посещаемых студентом в текущем семестре.
	8. Покупал ли студент в книжном магазине какие-либо предметы одежды?
	9. Способ оплаты покупки.

Определите, какие пункты опроса соответствуют категорийным переменным, а какие — числовым. Укажите уровень измерения.

**Ответы на контрольные задания**

1. Обратите внимание, что *уровень измерения* – синоним *шкалы измерения*
	1. Количество телефонов: случайная переменная – числовая, дискретная; шкала измерений – шкала отношений.
	2. Тип телефона: категорийная; номинальная.
	3. Количество разговоров: числовая, дискретная; шкала отношений.
	4. Продолжительность разговора: числовая, непрерывная; шкала отношений.
	5. Цвет телефона: категорийная; номинальная.
	6. Оплата: числовая, непрерывная; шкала отношений.
	7. Случайная переменная – числовая; шкала измерений – шкала отношений.
	8. Числовая; отношений.
	9. Числовая; отношений.
	10. Категорийная; номинальная.
	11. Категорийная; номинальная.
	12. Категорийная; номинальная.
	13. Числовая; отношений.
	14. Категорийная; номинальная.
	15. Категорийная; номинальная.
1. Используются материалы книги Левин и др. Статистика для менеджеров. – М.: Вильямс, 2005. – с. 52–57 [↑](#footnote-ref-1)