

Глава 4

Базовые (психо)метрики контента

Для того чтобы обрисовать современное представление о психометрическом моделировании, обратимся к метафоре. Представьте двух атлетов. Оба стараются изо всех сил, чтобы добиться желаемого результата. Это значит, что на результат оказывают влияние оба одновременно. Однако в результате поединка один одержит победу, а второй проиграет. Но кто из двух? Более сильный и ловкий, не так ли? Чтобы прогнозировать результат поединка, мы изучаем характеристику каждого атлета, статистику их побед и поражений, динамику массы тела и прошлых травм, отмечаем любимые приемы и слабые места, смену тренеров и команд, анализируем их предыдущие поединки, мысленно сравниваем друг с другом. А чтобы объяснить результаты текущего поединка, обращаем внимание на их взаимодействие, на условия поединка: кто оказался в новом для себя климате, а у кого произошла смена часовых поясов.



Теперь на место одного атлета поставим студента, а на место второго — задание. Студент пытается решить задачу. Но сможет ли? На результаты этого «поединка» влияют «оба атлета»: как студент, так и задача. Чем сильнее студент, тем выше шансы на верное решение. Чем сложнее задача, тем эти шансы ниже. Важнейшая задача психометрики — исследовать максимально подробно и точно обе стороны: характеристики и студента, и задачи.

В этой главе мы познакомимся с характеристиками задач в узком смысле и учебного контента — в широком. И разберем несколько метрик заданий по принципу от простого к сложному.

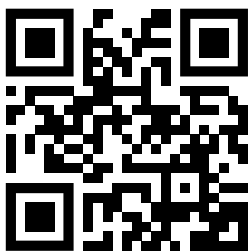
Начиная с этой главы, я буду приводить примеры кода для расчетов метрик на языке R. Поскольку части книги, включающие вычислительный код, рассчитаны на начинающих EdTech-аналитиков или методистов с базовым знанием языков программирования, объяснения будут подробно аннотированы.

Доля верных ответов при одной попытке

Начнем с самого простого случая. Студент выполняет задание, и у него только одна попытка. Можно считать, что у задания есть некое общее свойство, определяющее, насколько хорошо студенты с ним справятся. Обычно это свойство называют *трудностью задания*, и выражается оно по-разному.

Самый простой способ — рассчитать долю верных ответов. Для этого определяем долю верных ответов на это задание относительно общего количества ответов на него. Или, другими словами, вычисляем долю студентов, которые справились с этим заданием, от числа всех студентов,

приступивших к его решению. Этот способ пришел из классической психометрической теории (или классической теории тестирования). Доля верных ответов будет иметь границы от 0 до 1, где 0 означает, что с заданием не справился никто из студентов, и 1 – что с задачей справились все студенты. Как было отмечено, доля верных ответов часто интерпретируется как трудность задания. Но важно обратить внимание, что высокие значения соответствуют простым задачам, то есть чем выше доля верных ответов, тем легче задание.



Рассчитаем долю верных ответов на реальном примере. У нас есть датасет (скачайте этот файл под названием *book_dataset_1.tsv* по ссылке clck.ru/3EivRg и поместите в рабочую директорию R). Он содержит несколько переменных: `student_id`, `course_id`, `topic_id`, `lesson_id`, `task_id`, `is_correct`, но для расчета доли верных ответов в этом простейшем случае нужны только три: `student_id`, `task_id` и `is_correct`. Первая представляет анонимизированный идентификатор студента, вторая – идентификатор задания, а третья – маркировку ответа студента на задание, где 1 – ответ верный и 0 – ответ неверный. У студентов из этого примера была только одна попытка решения, поэтому в датасете только одна пара «студент – задание» с соответствующей этой паре маркировкой ответа.

Начнем с того, что подгрузим этот датасет с помощью следующей строчки кода:

```
df <- read.table("book_dataset_1.tsv", header =
TRUE, quote = "", sep = "\t")
```

Проще всего познакомиться со структурой и содержанием датасета с помощью команды которая всегда выводит первые шесть строк данных – `head(as.data.frame(df))`. Выполнив ее, вы получите следующее (скрин. 1):

	student_id	course_id	topic_id
1	d8800e0cba2cf3dce798f3dd4aec441f	50a9e052	9f44ae60
2	d8800e0cba2cf3dce798f3dd4aec441f	50a9e052	9f44ae60
3	d8800e0cba2cf3dce798f3dd4aec441f	50a9e052	9f44ae60
4	d8800e0cba2cf3dce798f3dd4aec441f	50a9e052	9f44ae60
5	d8800e0cba2cf3dce798f3dd4aec441f	50a9e052	9f44ae60
6	d8800e0cba2cf3dce798f3dd4aec441f	50a9e052	48b34158
	lesson_id	task_id	is_correct
1	3ada2f9f	d596380e58edaf48	0
2	3ada2f9f	205232a29ee3585f	1
3	3ada2f9f	a8597dc43bcc0e83	0
4	d59bafc3	7713445fd963da0d	1
5	d59bafc3	9cf4b9615e066f3a	1
6	09544a3d	bc80d975dd59f5c8	1

Скриншот 1

Мы видим идентификатор студента, идентификаторы курса, темы, урока и задания и маркировку ответа на задание. Мы понимаем, что студент с идентификатором `d8800e0cba2cf3dce798f3dd4aec441f` справился со вторым, четвертым, пятым и шестым заданиями и не справился с первым и вторым. Все шесть заданий относятся к курсу с идентификатором `50a9e052`, при этом первые пять – к теме с идентификатором `9f44ae60`, а шестое – к теме с идентификатором `48b34158`. Первые три задания относятся к уроку `3ada2f9f`, четвертое и пятое – к уроку `d59bafc3` и шестое – к уроку `09544a3d`. В этом продукте задания вложены в уроки, уроки в темы, темы в курсы, а курс составляет отдельный продукт (в этом датасете он один).

Далее нам нужно установить и загрузить пакет `plyr`. Сделать это можно последовательно с помощью команд

`install.packages("dplyr")` и `library(dplyr)`. Первую нужно использовать один раз, так как она скачивает необходимый пакет и устанавливает его на ваш компьютер. А вторую команду важно задеять каждый раз перед началом рабочей сессии (открыли R, воспользовались командой и далее применяем код из этой книги).

Итак, для того чтобы рассчитать долю верных ответов, необходимо воспользоваться следующим кодом:

```
p_i <- df %>%
  group_by(task_id) %>%
  summarise(
    n_Correct = sum(is_correct == 1),
    n_Incorrect = sum(is_correct == 0),
    p_i = round(n_Correct / (n_Correct +
      n_Incorrect), 2),
    .groups = "drop"
  ) %>%
  select(task_id, p_i)
```

Разберем этот код последовательно. Сначала данные `df` группируются по заданиям с помощью команды `group_by(task_id)`. Условно все ответы на отдельное задание раскладываются по корзинкам, каждая из которых соответствует одному заданию и маркируется его идентификатором. Затем в каждой корзинке считаем количество верных ответов (`n_Correct`) с помощью команды `n_Correct = sum(is_correct == 1)` и количество неверных ответов (`n_Incorrect`) с помощью команды `n_Incorrect = sum(is_correct == 0)`. После мы можем рассчитать долю верных ответов в каждой корзинке, то есть для каждого задания, с помощью команды `p_i = round`

$(n_Correct / (n_Correct + n_Incorrect), 2)$. Вы видите, что мы вычисляем долю студентов, справившихся с заданием верно ($n_Correct$), от общего числа студентов, решавших это задание и, соответственно, давших верный и неверный ответы ($n_Correct + n_Incorrect$). В результате получаем значение доли верных ответов (p_i) в каждой корзинке. Эти значения помещаются в таблицу с аналогичным названием (p_i), где p – proportion (proportion correct, доля верных ответов), а i – это индекс, которым обычно обозначают задания (от слова items). В таблице имеются два столбца: первый с идентификатором задания $task_id$, а второй – с соответствующей ему долей верных ответов p_i .

Рассмотрим иллюстрационную (а не рассчитанную на датасете) таблицу (табл. 9).

task_id	p_i	Мы видим, что значения доли верных ответов варьируются. С первым заданием с идентификатором t001 справились 12% студентов, со вторым – 97%, с третьим и четвертым заданиями – 35 и 81% соответственно. Как интерпретировать эти значения?
t001	0,12	
t002	0,97	
t003	0,35	
t004	0,81	

Таблица 9. Доли верных ответов для четырех заданий

С точки зрения учебных заданий оптимальными считаются те, доля верных решений в которых составляет 75–95%.

Наша практика показывает, что оптимальных учебных результатов (например, получают более высокие оценки за проектные работы и с наибольшей вероятностью завершают обучение) добиваются те студенты, у которых доля успехов в учебном процессе составляет 75–95%. Такую долю успеха формируют учебные задания с соответствующей долей верных решений

(при условии только одной попытки на их решение). Задания с долей верных решений в интервале от 50 до 75% считаются субоптимальными; с долей верных решений ниже 50% – фрустрирующими (или фрустрирующе трудными), а с долей верных решений выше 95% – скучными.

Эти градации я подсчитывал и валидировал во время работы в международном Практикуме (ныне TripleTen), поэтому их названия в оригинале были сформулированы на английском языке: frustrating (фрустрирующие, < 50%), suboptimal (субоптимальные, от 50 до 75%), optimal (оптимальные, от 75 до 95%) и boring (скучные, выше 95%). Отмечу, что в одном из современных исследований, опубликованном во влиятельном журнале Nature* в 2019 году, ученые из США эмпирически рассчитали оптимальную долю успехов и провалов в обучении: 85/15. Как видно, эта оптимальная категория находится ровно в середине нашей оптимальной категории (75–95%). Для удобства представим референсные значения в виде таблицы (табл. 10).

Референсный интервал	Интерпретация
Менее 0,50	Фрустрирующая (frustrating)
От 0,50 до 0,75	Субоптимальная (suboptimal)
От 0,75 до 0,95	Оптимальная (optimal)
Выше 0,95	Скучная (boring)

Таблица 10. Референсные интервалы с интерпретациями для доли верных ответов

* R. C. Wilson, A. Shenhav, M. Straccia & J. D. Cohen. The eighty five percent rule for optimal learning // Nature Communications. 2019. November 05. URL: <https://www.nature.com/articles/s41467-019-12552-4>.

Если же рассуждать не об учебных, а о проверочных заданиях или экзаменах, то существуют конвенциональные границы, предложенные в классической психометрической теории (или классической теории тестирования): задания с трудностью менее 0,1 считаются очень трудными (их верно решило менее 10% выборки); с трудностью от 0,1 до 0,2 – просто трудными; диапазон от 0,2 до 0,8 наиболее продуктивен для измерений с целью подведения итогов или отбора лучших; задания с трудностью выше 0,8 оцениваются как легкие, а выше 0,9 – очень легкие.

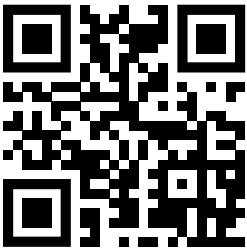
В целом нужно заметить, что оценка доли верных ответов обычно зависит от ситуации оценивания, типа заданий, от обучающейся аудитории – иными словами, от всего того контекста, в котором идет борьба студентов с заданиями. Тут важно сделать оговорку: если в спорте стремятся победить обе стороны, то в образовании наша задача все же сбалансировать взаимодействие студентов с заданиями, ориентируясь на пользу для студента.

Мини-практикум. С помощью команды `head(as.data.frame(p_i))` выведите первые шесть рассчитанных значений доли верных ответов. К каким категориям с точки зрения доли верных ответов относится каждое из этих шести заданий?

Правильный ответ: 0,90 (оптимальное), 0,55 (субоптимальное), 0,84 (оптимальное), 1,00 (скучное), 0,77 (оптимальное), 1,00 (скучное).

Доля верных ответов при условии нескольких попыток

А теперь усложним рассмотренный случай и представим, что у студентов есть возможность совершить несколько попыток. В этом случае в датасете каждая строка будет показывать результат одной попытки решения задания. В предыдущей главе мы говорили о том, что если студенту доступны несколько попыток для решения задания, то обязательным условием является добавление переменной `timestamp` (то есть сведений о времени выполнения действия; в приведенном примере эта переменная называется `timestamp`).



Соответственно, минимально достаточная структура датасета должна быть представлена уже четырьмя столбцами: `student_id`, `task_id`, `is_correct` и `timestamp`. В нашем втором датасете эти переменные есть (скачайте соответствующий датасет под названием

`book_dataset_2a.tsv` по ссылке clck.ru/3Eivwc и поместите в рабочую директорию R), но кроме них в датасете содержатся еще и переменные `course_id`, `topic_id` и `lesson_id`, показывающие иерархическую вложенность содержания.

Доля верных ответов в первой попытке

Есть несколько способов выразить свойство задания (условно — его трудность). Один из них — рассчитать долю верных ответов по первой попытке. Давайте сделаем это.



[Почитать описание, рецензии
и купить на сайте](#)

Лучшие цитаты из книг, бесплатные главы и новинки:

