

Анализ категориальных данных

Занятия 1 – 2. Модели бинарного выбора: спецификация, интерпретация

21 февраля 2020

Вопрос

Вспомните, что такое линейная вероятностная модель (linear probability model)

Вопрос

Вспомните, что такое линейная вероятностная модель (linear probability model)

Ответ

Это результат оценивания классической линейной регрессии применительно к случаю бинарного отклика:

$y_i = \beta_0 + \beta x_i + e_i$, где y_i принимает только два значения (к примеру, 1 – приняли рукопись к публикации, 0 – в противном случае)

В этом случае предсказанное значение отклика (\hat{y}_i) – это вероятность того, что Y принимает значение 1

$$E(y_i|x_i) = 1 \times P(y_i = 1|x_i) + 0 \times P(y_i = 0|x_i)$$

Вопрос

В чем основные ограничения линейной вероятностной модели?

Вопрос

В чем основные ограничения линейной вероятностной модели?

Ответ

- 1 Предсказанные значения отклика выходят за допустимые границы, может быть меньше 0 или больше 1 (при этом $\hat{y}_i = P(y_i = 1)$)
- 2 Содержательно не всегда правдоподобной является линейная взаимосвязь вероятности «успеха» и объясняющей переменной

Вопрос

Рассмотрим альтернативу. Объясните суть подхода, основанного на латентной зависимой переменной?

Вопрос

Рассмотрим альтернативу. Объясните суть подхода, основанного на латентной зависимой переменной?

Ответ

Вопрос

Рассмотрим альтернативу. Объясните суть подхода, основанного на латентной зависимой переменной?

Ответ

Мы допускаем, что существует некоторая ненаблюдаемая переменная y_i^* , принимающая любые значения $(-\infty; +\infty)$

Условно ее можно интерпретировать как склонность к «успеху» (склонность к тому, что наблюдаемый $y_i = 1$)

На основе значений y_i^* определяются значения исходного y_i .

Если $y_i^* > 0$, то $y_i = 1$

Если $y_i^* \leq 0$, то $y_i = 0$

Вопрос

Запишите спецификацию модели с y_i^* в качестве отклика. Как оценивается такая модель?

Вопрос

Запишите спецификацию модели с y_i^* в качестве отклика. Как оценивается такая модель?

Ответ

Важно, что латентная зависимая переменная линейным образом связана с объясняющими переменными:

$$y_i^* = \beta_0 + \beta x_i + e_i$$

Так как отклик ненаблюдаемый, нам нужны допущения о распределении ошибок:

- 1 $\epsilon \sim N(0, 1)$ (probit-model)
- 2 стандартное логистическое распределение $\epsilon \approx N(0, 3.29)$ (logit-model). $F(\epsilon) = \frac{\exp(\epsilon)}{1 + \exp(\epsilon)}$

Вопрос

Покажите, что $P(y_i = 1) = F(\beta_0 + \beta x_i)$ (F – функция распределения)

Вопрос

Покажите, что $P(y_i = 1) = F(\beta_0 + \beta x_i)$ (F – функция распределения)

Ответ

$P(y_i = 1) = P(y_i^* > 0) = P(\beta_0 + \beta x_i + e_i > 0) = P(e_i \leq \beta_0 + \beta x_i)$,
а функция распределения – это и есть вероятность того, что сл. величина не превышает указанное значение.

К примеру, для логит-модели:

$$P(y_i = 1) = F(\beta_0 + \beta x_i) = \frac{\exp(\beta_0 + \beta x_i)}{1 + \exp(\beta_0 + \beta x_i)}$$

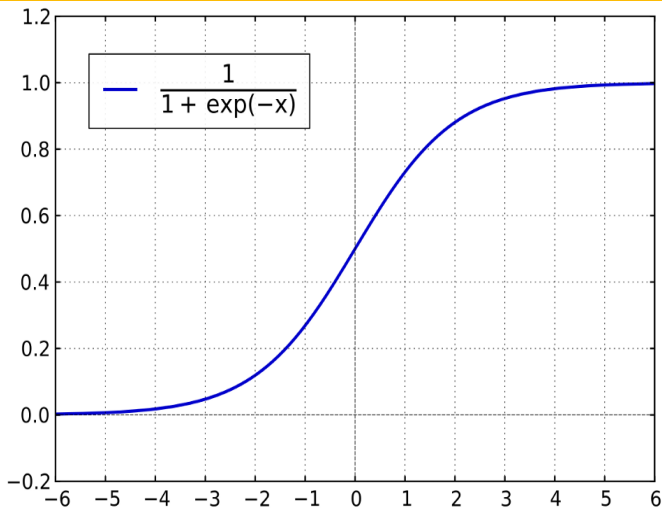
Можно обойтись и без латентного y_i^* :

Можно обойтись и без латентного y_i^* :

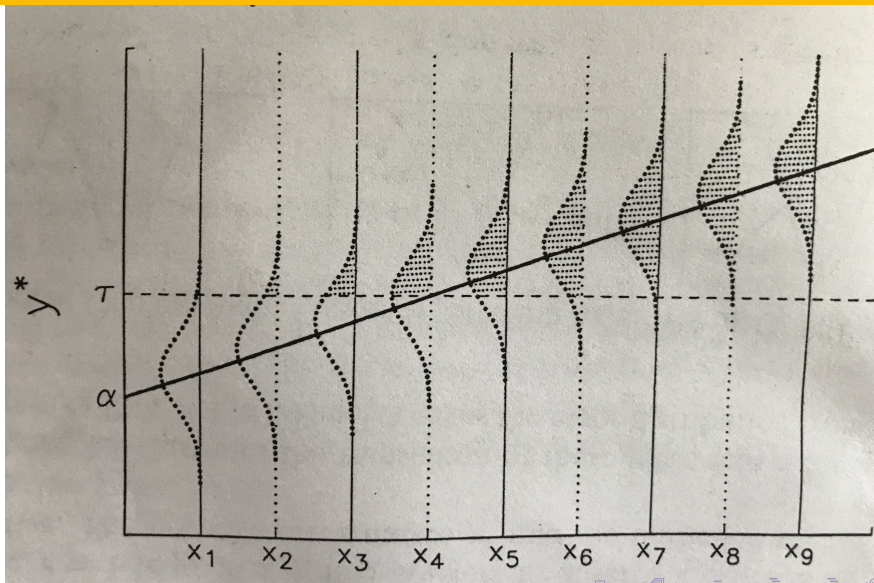
Ответ

1. Перейдем от $P(y_i = 1)$ к шансам $\frac{P(y_i = 1)}{1 - P(y_i = 1)}$
2. Запишем $P(y_i = 1)$ как функцию распределения:
$$\frac{\exp(\beta_0 + \beta x_i)}{1 + \exp(\beta_0 + \beta x_i)} = \frac{P(y_i = 1)}{1 - P(y_i = 1)}$$
3. $\ln\left(\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right) = \beta_0 + \beta x_i$ (логит линейным образом связан с объясняющими переменными)

Зависимость $P(Y = 1)$ от X ...



...в результате той самой ползущей «улитки»



Вопрос: каким образом интерпретируются оценки коэффициентов в логистических моделях?

Вопрос: каким образом интерпретируются оценки коэффициентов в логистических моделях?

Ответ

1. Предельные эффекты (для непрерывных переменных)

Вопрос: каким образом интерпретируются оценки коэффициентов в логистических моделях?

Ответ

- 1 Предельные эффекты (для непрерывных переменных)
- 2 Дискретное изменение в вероятности

Вопрос: каким образом интерпретируются оценки коэффициентов в логистических моделях?

Ответ

- 1 Предельные эффекты (для непрерывных переменных)
- 2 Дискретное изменение в вероятности
- 3 Отношения шансов (odds ratio)

Задача в преддверии праздников



Задача в преддверии праздников

Ниже представлены оценки логит-модели, предсказывающей вероятность того, что девушка получит подарок на 8 марта от своего молодого человека, на основании следующих предикторов:

- `present23Feb` – девушка поздравила своего МЧ с 23 февраля (дамми-переменная: 1 – поздравила, 0 – нет)
- `alittleforgetful` – ее МЧ слегка забывчив (1 – да, 0 – нет)

Оценки модели

$$0.4 + 0.7 \times \text{present23Feb}_i - 2.1 \times \text{alittleforgetful}_i$$

Задача в преддверии праздников

Ниже представлены оценки логит-модели, предсказывающей вероятность того, что девушка получит подарок на 8 марта от своего молодого человека, на основании следующих предикторов:

- `present23Feb` – девушка поздравила своего МЧ с 23 февраля (дамми-переменная: 1 – поздравила, 0 – нет)
- `alittleforgetful` – ее МЧ слегка забывчив (1 – да, 0 – нет)

Оценки модели

$$0.4 + 0.7 \times \text{present23Feb}_i - 2.1 \times \text{alittleforgetful}_i$$

Задание

Проинтерпретируйте оценки в терминах отношений шансов