Регрессионный анализ социально-экономических процессов

Модели бинарного выбора

26 января 2026

Вопрос

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

Вопрос

Покажем, 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, где, к примеру, 1 – приняли рукопись к публикации, 0 – в противном случае.

Вопрос

Покажем, 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, где, 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:
$E(y_i|x_i) = 1 \times P(y_i = 1|x_i) + 0 \times P(y_i = 0|x_i) = P(y_i = 1|x_i)$

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

Вопрос

В чем основные ограничения линейной вероятностной модели?

Ответ

1. Предсказанные значения отклика выходят за допустимые границы, может быть меньше 0 или больше 1
2. Содержательно не всегда правдоподобной является линейная взаимосвязь вероятности «успеха» и объясняющей переменной

Вопрос

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

Вопрос

Рассмотрим альтернативу. В чем 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

Запишем спецификацию модели с y_i^* в качестве отклика. Какие допущения делаем об ошибках?

Вопрос

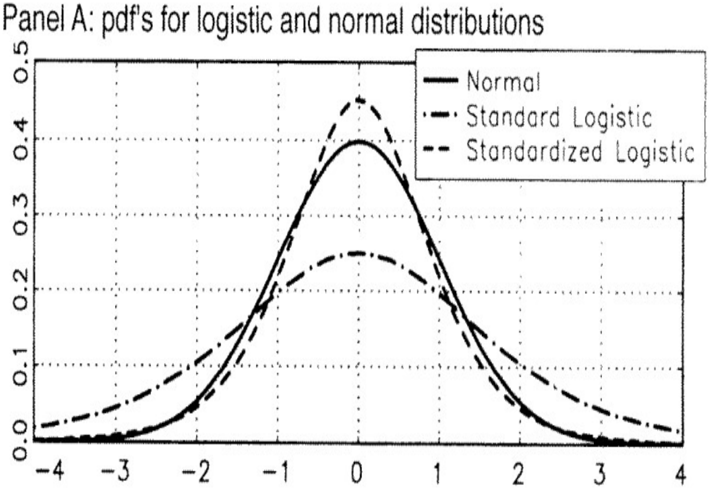
Запишем спецификацию модели с y_i^* в качестве отклика. Какие допущения делаем об ошибках?

Ответ

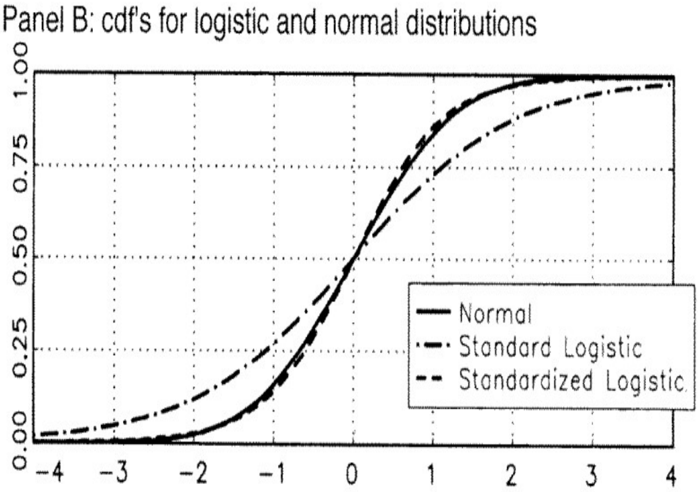
Важно, что латентная зависимая переменная линейным образом связана с объясняющими переменными:
$y_i^* = \beta_0 + \beta x_i + e_i$
Так как отклик ненаблюдаемый, нам нужны допущения о распределении ошибок:

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

Графики функций плотности



Графики функций распределения



Вопрос

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

Вопрос

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

Ответ

$P(y_i = 1) =$

Вопрос

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

Ответ

$P(y_i = 1) = P(y_i^* > 0) =$

Вопрос

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

Ответ

$P(y_i = 1) = P(y_i^* > 0) = P(\beta_0 + \beta x_i + e_i > 0) =$

Вопрос

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

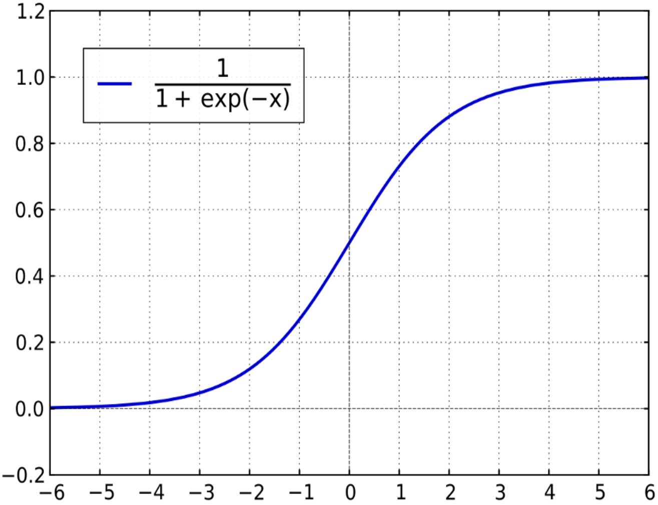
Ответ

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

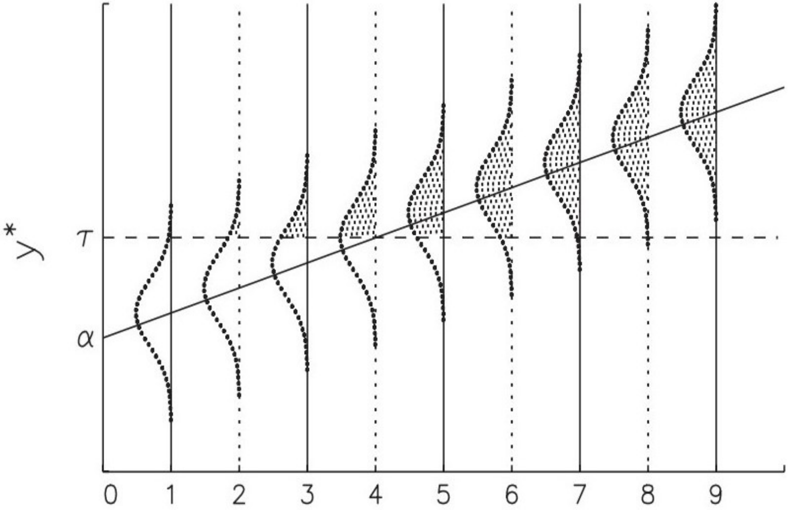
К примеру, для логит-модели:

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

Зависимость $P(Y = 1)$ от X ...



…в результате той самой ползущей «улитки»



Можно обойтись и без латентного y_i^*:

Можно обойтись и без латентного y_i^*:

Ответ

1. Перейдем от $P(y_i = 1)$ к шансам $\dfrac{P(y_i = 1)}{1 - P(y_i = 1)}$

Можно обойтись и без латентного y_i^*:

Ответ

1. Перейдем от $P(y_i = 1)$ к шансам $\dfrac{P(y_i = 1)}{1 - P(y_i = 1)}$
2. Запишем $P(y_i = 1)$ как функцию распределения:
$$\frac{\dfrac{exp(\beta_0 + \beta x_i)}{1 + exp(\beta_0 + \beta x_i)}}{1 - \dfrac{exp(\beta_0 + \beta x_i)}{1 + exp(\beta_0 + \beta x_i)}} = exp(\beta_0 + \beta x_i)$$

Можно обойтись и без латентного y_i^*:

Ответ

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

Вопрос

Что из себя представляет confusion matrix? Как ее построить?

Вопрос

Что из себя представляет confusion matrix? Как ее построить?

Ответ

1. Сначала нужно сохранить предсказанные моделью вероятности $P(Y = 1)$
2. Далее выбрать порог отсечения: к примеру, если $P(Y = 1)$ более 0.5 отнести наблюдение к классу 1, и в противном случае – к классу 0.
3. Далее на основе предсказанных и наблюдаемых значений можно построить аналог таблицы сопряженности

Вопрос

Рассмотрим элементы confusion matrix подробнее.

Вопрос

Рассмотрим элементы confusion matrix подробнее.

Ответ

$$\left(\begin{array}{ccc} & \color{red}{data: Y=1} & \color{red}{Y=0} \\ \color{blue}{prediction: Y=1} & TP & FP \\ \color{blue}{prediction: Y=0} & FN & TN \end{array}\right), \text{ где}$$

TP – истинно «положительные» значения (в реальности относится к классу 1 и классифицировано моделью так же)

TN – по аналогии: истинно «отрицательные» значения

FP – допущена ошибка классификатором: отнесли к классу 1 («положительные»), а на самом деле – класс 0

FN – допущена ошибка классификатором: отнесли к классу 0 («отрицательные»), а на самом деле – класс 1

Вопрос

Определите по этой confusion matrix ошибку I рода, ошибку II рода и мощность критерия.

Вопрос

Определите по этой confusion matrix ошибку I рода, ошибку II рода и мощность критерия.

Ответ

$$\begin{pmatrix} & \color{red}{data: Y=1} & \color{red}{Y=0} \\ \color{blue}{prediction: Y=1} & TP & FP \\ \color{blue}{prediction: Y=0} & FN & TN \end{pmatrix},$$

ошибка I рода $= P(reject|H0) = \dfrac{FP}{FP+TN}$

ошибка II рода $= P(NOTreject|H1) = \dfrac{FN}{FN+TP}$

мощность $= P(reject|H1) = \dfrac{TP}{FN+TP}$

Чтобы confusion matrix не смогла Вас confuse:

1. Когда считаете ошибку I рода, вспоминайте, что теперь массив сужается только до H0 (класс 0 по ИСХОДНЫМ ДАННЫМ): отвержение при условии верной H0. Мысленно оставляйте в матрице только тот столбец, который соответствует (data: Y = 1), то есть, TN + FP. А дальше зададимся вопросом, когда допускается ошибка? FP – это, конечно, же ошибка, поэтому и получаем $\dfrac{FP}{FP+TN}$
2. По аналогии делайте и при расчете ошибки II рода: только теперь Вас интересует подмассив «класс 1»

Вопрос

Что такое меры чувствительности (sensitivity) и специфичности (specificity)?

Вопрос

Что такое меры чувствительности (sensitivity) и специфичности (specificity)?

Ответ

Когда считаем эти меры, нас всегда будет интересовать, какую долю наблюдений мы классифицировали моделью ВЕРНО (относительно исходных данных). Осталось только запомнить, что чувствительность – это про верные «положительные» наблюдения, а специфичность – про верные «отрицательные».

$$\text{Sensitivity} = \frac{TP}{TP+FN}; \text{Specificity} = \frac{TN}{TN+FP}$$

Несложно заметить, что

Sensitivity – это мощность критерия (которую мы всегда хотим максимизировать).
Specificity – это (1 – ошибка I рода), эту величину тоже хочется максимизировать.
Однако одновременно это сделать на практике сложно, для того, чтобы найти подходящее пороговое значение (насколько это возможно, максимизирующее мощность и минимизирующее ошибку I рода) нам пригодится ROC. См. полезный интерактив с бегунком по ROC – здесь.

Меры качества модели: R^2

Для логистических моделей, так же как и для классических линейных, существуют R^2, только они псевдо-R^2. Они основаны на функции правдоподобия модели и НЕ могут интерпретироваться как доля объясненной вариации. Подробнее про разные варианты pseudo-R^2 можно посмотреть здесь.