

Демонстрационная версия проверочной работы №2

Задание 1. Исследователь изучает государственные перевороты. Наблюдениями являются попытки государственных переворотов в мире с 1946 по 2012 год. Была оценена модель, в которой

y — данная попытка переворота увенчалась успехом ($y = 1$) или не увенчалась ($y = 0$),

$x^{(1)}$ — рост ВВП в последний год в этой стране, в %,

$x^{(2)}$ — коэффициент неравенства по доходам (джини) в этой стране, от 0 до 100,

$x^{(3)}$ — идёт ли в год попытки переворота Холодная Война (1 — да, 0 — нет),

1. Запишите в общем виде спецификацию указанной логистической регрессионной модели
2. Рассчитайте вероятность того, что попытка переворота в стране с падением ВВП на 5%, в условиях коэффициента неравенства, равного 30, и идущей Холодной Войны будет успешной, предположив стандартное логистическое распределение ошибок и зная, что $\hat{\beta}_0 = -1.244$, $\hat{\beta}_1 = -0.031$, $\hat{\beta}_2 = 0.011$, $\hat{\beta}_3 = 0.617$
3. Рассчитайте отношение шансов успеха попытки переворота, рассматривая при прочих равных условиях рост ВВП на 4 % и коэффициент неравенства по доходам равным 20 по сравнению с ростом ВВП на 5 % и коэффициентом неравенства по доходам равным 28

Задание 2.

1. **Отметьте ВСЕ верные утверждения, если таковые имеются. Если верных утверждений нет, то напишите в ответе «НЕТ» (1 балл)**
 - (a) F-тест используется для сравнения вложенных линейных и логистических регрессионных моделей
 - (b) С ростом меры чувствительности (sensitivity/recall) в контексте логистической регрессии мера специфичности будет увеличиваться
 - (c) R^2 — это квадрат косинуса угла между центрированными векторами y наблюдаемого и \hat{y} (предсказанного)
 - (d) Сравняются две вложенные регрессионные модели. Тест отношения правдоподобия показывает, что нулевая гипотеза отвергается, это означает, что нужно выбрать модель менее экономную
2. **Ответьте на вопросы ниже:**
 - 1) По результатам сравнения двух регрессионных моделей m_1 и m_2 были получены следующие значения информационного критерия AIC: $AIC_{m_1} = 1239.665$ и $AIC_{m_2} = 1292.047$. Объясните своими словами, что показывает AIC. Сделайте выбор: какая модель лучше с точки зрения указанных значений AIC? Свой ответ поясните
 - 2) Покажите, что предсказанная вероятность получить $Y = 1$ в логистической регрессионной модели бинарного выбора — это функция распределения от $X\hat{\beta}$

Задание 3. На основе представленной ниже confusion matrix, построенной по результатам оценивания модели с порядковым откликом, имеющим 3 категории:

Prediction	Reference		
	class1	class2	class3
class1	20	19	10
class2	71	130	32
class3	54	37	160

1. рассчитайте Accurasy и сравните это значение с Baseline Accurasy.
2. найдите значение чувствительности и специфичности для класса 3 и объясните своими словами, что данные меры показывают

Задание 4. Ниже представлена выдача по результатам оценивания логистической регрессионной модели с порядковым откликом (всего 4 категории от 1 – «полное несогласие» с утверждением до 4 – «полное согласие» с утверждением). Переменная male – дамми (1 – мужчина, 0 – женщина), переменная yr15 – дамми-переменная (принимает 1, если год опроса – 2015, 0 – 2013 год), age – возраст

Coefficients:

	Coef	Std. Error	t value
yr15	0.60865	0.078780	7.726
male	-0.71315	0.078125	-9.128
age	-0.02535	0.002321	-10.921

Thresholds:

	Coef	Std. Error	t value
1 2	-3.2783	0.1424	-23.0220
2 3	-1.4713	0.1270	-11.5893
3 4	0.3943	0.1238	3.1853

1. Рассчитайте вероятность попасть в категорию 3 («согласие с утверждением») зависимой переменной, если респондент – мужчина 35 лет, волна опроса – 2015 г.
2. Рассчитайте при условии тех же характеристик, что и в первом пункте, функцию распределения для категории 2
3. Дайте интерпретацию коэффициента при yr15 в терминах отношения шансов, проиллюстрировав идею о пропорциональности шансов
4. Для проверки допущения о пропорциональности шансов была оценена соответствующая мультиномиальная регрессионная модель. Как по оценкам этой модели можно определить, соблюдается ли данное допущение, есть ли смысл выбрать мультиномиальную модель вместо модели с порядковым откликом в допущении о пропорциональности шансов?