

Отчет по ДЗ 2

Добин Илья

БЭАД223

Задание 1

Пункт (а): Logit-модель бинарного выбора

$$P(\text{crime86}_i = 1 | X_i) = \Lambda(X_i' \beta) = \frac{\exp(X_i' \beta)}{1 + \exp(X_i' \beta)} \quad (1)$$

где X_i - вектор объясняющих переменных для индивида i , β - вектор коэффициентов.

Результаты оценки модели:

Logit Regression Results						
Dep. Variable:	crime86	No. Observations:	1500			
Model:	Logit	Df Residuals:	1489			
Method:	MLE	Df Model:	10			
Date:	Mon, 28 Apr 2025	Pseudo R-squ.:	0.07636			
Time:	19:51:08	Log-Likelihood:	-735.81			
converged:	False	LL-Null:	-796.64			
Covariance Type:	nonrobust	LLR p-value:	2.330e-21			
	coef	std err	z	P> z	[0.025	0.975]
const	-0.8840	0.167	-5.278	0.000	-1.212	-0.556
pcnv	-0.5634	0.171	-3.301	0.001	-0.898	-0.229
avgsen	3.8787	4.05e+04	9.57e-05	1.000	-7.94e+04	7.94e+04
totttime	-3.8869	4.05e+04	-9.59e-05	1.000	-7.94e+04	7.94e+04
ptime86	-0.0788	0.041	-1.921	0.055	-0.159	0.002
qemp86	0.0645	0.065	0.990	0.322	-0.063	0.192
inc86	-0.0114	0.002	-5.684	0.000	-0.015	-0.007
durat	0.0170	0.016	1.063	0.288	-0.014	0.048
black	0.7171	0.173	4.149	0.000	0.378	1.056
hispan	0.3259	0.157	2.071	0.038	0.017	0.634
born60	0.0179	0.149	0.120	0.904	-0.275	0.310

Рисунок 1. Logit

$$\begin{aligned} \log \left(\frac{P(\text{crime86} = 1)}{1 - P(\text{crime86} = 1)} \right) = & -0.884 - 0.5634 \cdot \text{pcnv} - 3.8787 \cdot \text{avgsen} \\ & - 3.8869 \cdot \text{totttime} - 0.0788 \cdot \text{ptime86} - 0.0645 \cdot \text{qemp86} \\ & + 0.0114 \cdot \text{inc86} + 0.017 \cdot \text{durat} + 0.07171 \cdot \text{black} \\ & + 0.3259 \cdot \text{hispan} + 0.0179 \cdot \text{born60} \end{aligned}$$

Пункт (b): Прогноз вероятностей

Прогнозные вероятности рассчитываются по формуле (1) Тогда вероятность не быть арестованным:

$$P(\text{crime86} = 0) = 1 - P(\text{crime86} = 1) = \frac{1}{1 + e^{X_i \beta}}$$

Пункт (с): Предельные эффекты

Для непрерывных переменных:

$$AME_k = \frac{1}{N} \sum_{i=1}^N \frac{\partial P(\text{crime86} = 1 | x_i)}{\partial x_{ik}} = \frac{1}{N} \sum_{i=1}^N \beta_k \cdot P_i \cdot (1 - P_i),$$

Для бинарных переменных (например, black):

$$AME_k = \frac{1}{N} \sum_{i=1}^N [P(\text{crime86} = 1 | x_{ik} = 1, \mathbf{x}_{i,-k}) - P(\text{crime86} = 1 | x_{ik} = 0, \mathbf{x}_{i,-k})],$$

Результаты расчета предельных эффектов:

Logit Marginal Effects						
=====						
Dep. Variable:	crime86					
Method:	dydx					
At:	overall					
=====						
	dy/dx	std err	z	P> z	[0.025	0.975]

pcnv	-0.0903	0.027	-3.333	0.001	-0.143	-0.037
avgsen	0.6215	6492.112	9.57e-05	1.000	-1.27e+04	1.27e+04
totttime	-0.6228	6492.112	-9.59e-05	1.000	-1.27e+04	1.27e+04
ptime86	-0.0126	0.007	-1.927	0.054	-0.025	0.000
qemp86	0.0103	0.010	0.991	0.322	-0.010	0.031
inc86	-0.0018	0.000	-5.791	0.000	-0.002	-0.001
durat	0.0027	0.003	1.065	0.287	-0.002	0.008
black	0.1149	0.027	4.234	0.000	0.062	0.168
hispan	0.0522	0.025	2.079	0.038	0.003	0.101
born60	0.0029	0.024	0.120	0.904	-0.044	0.050

Рисунок 2. Marginal Effects

Значимые эффекты ($p < 0.05$):

- **pcnv** ($AME = -0.0903^{***}$):

Увеличение доли предыдущих арестов, приведших к осуждению, на 1 единицу **снижает** вероятность ареста на 9.03%.

Интерпретация: Эффект сдерживания - высокая вероятность осуждения предотвращает преступления.

- **inc86** ($AME = -0.0018^{***}$):

Рост легального дохода на \$100 **снижает** вероятность ареста на 0.18%.

Интерпретация: Экономическая стабильность уменьшает мотивацию к преступлениям.

- **black** ($AME = 0.1149^{***}$):

Афроамериканцы имеют вероятность ареста на 11.49% **выше**.

- **hispan** (AME = 0.0522**):

Латиноамериканцы имеют вероятность ареста на 5.22% **выше**.

Незначимые эффекты ($p > 0.05$)

- **avgsen** и **tottime**:

Средняя длительность приговоров и общее время в тюрьме **не влияют** на вероятность ареста ($p = 1.000$).

- **ptime86** (AME = -0.0126, $p = 0.054$):

Пограничная значимость: каждый месяц в тюрьме в 1986 г. может снижать вероятность ареста на 1.26%.

- **qemp86, durat, born60**:

Занятость, длительность безработицы и год рождения **не оказывают** значимого влияния.

Пункт (d): Отношение шансов

Отношение шансов для переменной x_j :

$$OR_j = \exp(\hat{\beta}_j) \quad (2)$$

Результаты:

	OR	2.5%	97.5%	p-value
const	0.413126	0.297525	0.573644	0.0000
pcnv	0.569257	0.407412	0.795395	0.0010
avgsen	48.359621	0.000000	inf	0.9999
tottime	0.020510	0.000000	inf	0.9999
ptime86	0.924231	0.852824	1.001617	0.0548
qemp86	1.066650	0.938796	1.211917	0.3219
inc86	0.988648	0.984764	0.992548	0.0000
durat	1.017159	0.985751	1.049568	0.2877
black	2.048402	1.459805	2.874322	0.0000
hispan	1.385259	1.017624	1.885710	0.0384
born60	1.018106	0.759902	1.364043	0.9043

Рисунок 3. Отношения шансов

- **const (OR = 0.413, $p < 0.001$)**: Базовый уровень вероятности ареста значительно ниже среднего.
- **pcnv (OR = 0.569, $p = 0.001$)**: Каждое увеличение доли предыдущих осуждений на 1 единицу снижает шансы ареста на 43.1%. Это указывает на сильный сдерживающий эффект судебной системы.

- **avgsen (OR = 48.360, p = 0.999) и tottime (OR = 0.021, p = 0.999):** Оба показателя, связанные с длительностью наказаний, статистически незначимы. Широкие доверительные интервалы (включая бесконечность) свидетельствуют о проблемах в оценке этих параметров.
- **ptime86 (OR = 0.924, p = 0.055):** Каждый дополнительный месяц, проведённый в тюрьме в 1986 году, снижает шансы ареста на 7.6%, но этот эффект имеет лишь пограничную статистическую значимость.
- **inc86 (OR = 0.989, p < 0.001):** Каждые дополнительные \$100 легального дохода уменьшают вероятность ареста на 1.1%, что подтверждает важность экономических факторов в предотвращении преступности.
- **black (OR = 2.048, p < 0.001):** Афроамериканцы имеют в 2.05 раза более высокие шансы ареста по сравнению с другими группами после контроля других факторов.
- **hispan (OR = 1.385, p = 0.038):** Латиноамериканцы имеют на 38.5% более высокую вероятность ареста, хотя этот эффект менее выражен, чем у афроамериканцев.

Пункт (е): Различие вероятностей между группами

Вероятность ареста для афроамериканца: 0.2671

Вероятность ареста для латиноамериканца: 0.1977

Средняя разница вероятностей (P): 0.0694

Используем Z-тест для проверки значимости разности вероятностей:

$$Z = \frac{\Delta P}{SE(\Delta P)}, \quad \text{где SE} — \text{стандартная ошибка разности.}$$

Z-статистика: 1.923, p-value: 0.0545 \Rightarrow отвергаем $H_0 : \Delta P = 0$

Пункт (f): Ordered Logit модель

OrderedModel Results						
Dep. Variable:	narr86		Log-Likelihood:		-1034.7	
Model:	OrderedModel		AIC:		2107.	
Method:	Maximum Likelihood		BIC:		2208.	
Date:	Mon, 28 Apr 2025					
Time:	19:51:11					
No. Observations:	1500					
Df Residuals:	1481					
Df Model:	10					
	coef	std err	z	P> z	[0.025	0.975]
pcnv	-0.5150	0.169	-3.056	0.002	-0.845	-0.185
avgse	1.9888	124.150	0.016	0.987	-241.340	245.318
totttime	-2.0006	124.150	-0.016	0.987	-245.330	241.329
ptime86	-0.0788	0.040	-1.951	0.051	-0.158	0.000
qemp86	0.0348	0.064	0.542	0.588	-0.091	0.161
inc86	-0.0112	0.002	-5.607	0.000	-0.015	-0.007
durat	0.0141	0.015	0.909	0.363	-0.016	0.044
black	0.7349	0.169	4.336	0.000	0.403	1.067
hispan	0.3336	0.156	2.141	0.032	0.028	0.639
born60	-0.0094	0.148	-0.064	0.949	-0.299	0.280
0/1	0.8310	0.165	5.041	0.000	0.508	1.154
1/2	0.4986	0.063	7.960	0.000	0.376	0.621
2/3	-0.1257	0.146	-0.863	0.388	-0.411	0.160
3/4	-0.3771	0.238	-1.584	0.113	-0.844	0.090
4/5	-1.1673	0.448	-2.607	0.009	-2.045	-0.290
5/6	-0.1540	0.363	-0.425	0.671	-0.865	0.557
6/7	-0.3592	0.588	-0.611	0.541	-1.511	0.792
7/9	-0.8986	1.006	-0.893	0.372	-2.870	1.073
9/10	-0.3633	1.019	-0.357	0.721	-2.361	1.634

Рисунок 4. Ordered Logit

$$\log \left(\frac{P(narr86 \leq j)}{P(narr86 > j)} \right) = \alpha_j - (X\beta), \quad j = 0, 1, 2, \dots$$

где:

- α_j — пороговые значения (cutoffs),
- $X\beta = \beta_0 + \beta_1 \cdot pcnv + \beta_2 \cdot avgse + \dots$

лень фул выписывать лол

Пункт (g): Прогноз вероятностей для ordered logit

В Ordered Logit-модели вероятность того, что $narr86 = 0$ (ни одного ареста), вычисляется как:

$$P(narr86 = 0) = \frac{e^{\alpha_0 - X\beta}}{1 + e^{\alpha_0 - X\beta}}$$

где: - α_0 — первый порог из результатов модели

Пункт (h): Проверка гипотезы параллельности

Проверяемая гипотеза:

$$H_0 : \text{Коэффициенты } \beta \text{ одинаковы для всех пороговых уровней.}$$

H_1 : Коэффициенты различаются (модель не соответствует предположению).

Сравнив коэффициенты визуально, видим, что они отличаются. Но для формального вывода рассчитаем статистику Брана через LR-тест LR-статистика: 18.755, p-value: 0.4727.
 $p\text{-value} > 0.05 \Rightarrow H_0$ не отвергается \Rightarrow модель корректна