

Семинар 29.

1. В задаче используются данные (Mroz, 1975). Пуассоновская регрессия для моделирования количества детей в семье:

$$P(Nkids = y_i) = \exp^{-\lambda_i} \frac{\lambda_i^{y_i}}{y_i!},$$

где $\lambda_i = \exp(\beta_1 + \beta_2 AGE_i + \beta_3 AGE_i^2 + \beta_4 WE_i + \beta_5 INCOME_i)$,

y — количество детей в семье,

AGE — возраст женщины (в годах),

AGE^2 — квадрат возраста женщины,

WE — образование женщины (в годах),

$INCOME$ — доход семьи в \$10000.

Ниже в таблице приведены результаты оценивания методом максимального правдоподобия.

+-----+ Poisson Regression Dependent variable NKIDS Number of observations 753 Iterations completed 7 Log likelihood function -1083.397 Number of parameters 5 Restricted log likelihood -1279.522 +-----+					
Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
Constant	-7.64180956	1.14268278	-6.688	.0000	
AGE	.49624655	.05663388	8.762	.0000	42.5378486
AGE2	-.00686403	.00069963	-9.811	.0000	1874.54847
WE	-.03430021	.01448182	-2.369	.0179	12.2868526
INCOME	.01193400	.02569902	.464	.6424	2.30805950
+-----+					
Matrix Cov.Mat. has 5 rows and 5 columns.					
	1	2	3	4	5
1	1.30572	-.06373	.00078	-.00319	.00284
2	-.06373	.00321	-.3948059D-04	.3794861D-04	-.00012
3	.00078	-.3948059D-04	.4894781D-06	-.3460068D-06	.1216040D-05
4	-.00319	.3794861D-04	-.3460068D-06	.00021	-.00014
5	.00284	-.00012	.1216040D-05	-.00014	.00066
+-----+					

- Оцените эффект увеличения возраста на 1 год на среднее (expected) количество детей.
- Покажите, что выборочное среднее оценок $\hat{\lambda}_i$ равно выборочному среднему y_i .
- Протестируйте на 5% уровне значимости гипотезу о совместной незначимости всех регрессоров AGE , AGE^2 , WE , $INCOME$ при помощи теста отношения правдоподобия (LR-тест).
- Укажите ограничения Пуассоновской регрессии. Какие модели Вы можете предложить для преодоления этих ограничений.

2. Набор данных *crime.dta* содержит данные по арестам в течение 1986 года по 2725 мужчин, родившимся в Калифорнии в 1960 или 1961 гг. Каждый из мужчин в выборке был арестован по крайней мере однажды до 1986 г. Вопрос: что объясняет, что мужчина снова был арестован в течение 1986 г. (сколько раз, и т.д.). Имеются следующие данные:

narr86 — количество арестов в 1986 г.

pcnv — доля предыдущих арестов, приведших к осуждению (прокси для неотвратимости наказания)

avgsen — среднее время срока заключения по предыдущим случаям осуждения (в месяцах) (прокси для суровости наказания)

tottime — общее время, проведенное в тюрьме после достижения возраста 18 лет (в месяцах)

ptime86 — число месяцев в тюрьме во время 1986 г. (не может быть арестован, пока в тюрьме)

qemp86 — количество кварталов, в которых имел работу, в течение 1986 г. (возможности на рынке труда)

inc86 — легальный доход в 1986 г., \$100

durat — длительность последнего периода безработицы (в месяцах)

black =1, если афроамериканец

hispan =1, если латиноамериканец

born60 =1, если родился в 1960 г.

crime86 =1, если был арестован хотя бы однажды в 1986 г. ($\text{genr crime } 86 = \text{narr86} > 0$)

Количество арестов моделируется с помощью пуассоновской регрессии: результаты оценивания которой представлены ниже:

Dependent Variable: NARR86				MODEL 3
Method: ML/QML - Poisson Count (Quadratic hill climbing)				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.616217	0.063618	-9.686227	0.0000
PCNV	-0.401797	0.084783	-4.739146	0.0000
AVGSEN	0.005603	0.007437	0.753388	0.4512
PTIME86	-0.093507	0.020314	-4.603135	0.0000
QEMP86	-0.037305	0.028907	-1.290529	0.1969
INC86	-0.008127	0.001037	-7.833997	0.0000
BLACK	0.661036	0.073871	8.948584	0.0000
HISPAN	0.501855	0.073884	6.792448	0.0000
R-squared	0.075108	Mean dependent var		0.404404
Adjusted R-squared	0.072725	S.D. dependent var		0.859077
S.E. of regression	0.827249	Akaike info criterion		1.657495
Sum squared resid	1859.353	Schwarz criterion		1.674846
Log likelihood	-2250.337	Hannan-Quinn criter.		1.663767
Restr. log likelihood	-2441.921	Avg. log likelihood		-0.825812
LR statistic (7 df)	383.1683	LR index (Pseudo-R2)		0.078456
Probability(LR stat)	0.000000			

- (а) С помощью модели пуассоновской регрессии оцените вероятность быть

неарестованным (P_0) для мужчины со следующими характеристиками:
 $avgseu = 0$, $inc86 = 50$, $pcnv = 0.1$, $ptime86 = 0$, $qemp86 = 3$, $black = 1$.

- (b) Для мужчины из пункта (a) рассчитайте предельный эффект от дохода в 1986 года для вероятности быть неарестованным.
- (c) Проинтерпретируйте результаты оценивания модели.