ДЗ-03 РЕШЕНИЕ

Problem 1

CHOICE Data. In the UK, an important career choice is made at age 16. At this age, all children sit national exams. A few months later, they have to decide whether to stay at school or to leave full-time education. If they leave, they can choose between a regular job, or some type of apprenticeship, combining education with work. In this set of exercises, we will examine which factors determine this choice.

We will use data from the UK National Child and Development Survey. This data set covers individuals born in the UK in March 1958. See Micklewright (1986) for a detailed description of this data source. Data on these respondents are collected at various stages of their life cycle. We use a subsample of boys and girls, excluding those living in Scotland. Most of the variables we use are measured at age 16.

At16	Continuation decision taken at age 16: 1: stays at school, 2: apprenticeship, 3: regular job.
Able7	General ability test score, measured at age 7.
Loginc	Log family income (at age 16)
Ctratio	Number of children per teacher at school level (school quality indicator)
Oldsib	Number of older siblings (at age 16)
Yngsib	Number of younger siblings (at age 16)
Etot	Number of O-levels obtained at national exams at age 16 (prior to continuation decision)
Female	1 for girls, 0 for boys

sum school able7 loginc ctratio oldsib yngsib etot female

Variable	Obs	Mean	Std. Dev.	Min	Max
school	3,423	.3131756	.4638529	0	1
able7	3,423	73.93441	20.44682	5	100
loginc	3,423	3.863004	.3924542	.9162908	5.105946
ctratio	3,423	17.11922	1.973063	4.8	35.28571
oldsib	3,423	.4271107	.6422592	0	8
yngsib	3,423	1.203623	1.24346	0	9
etot	3,423	6.127666	3.346675	0	22
female	3,423	.4995618	.5000729	0	1

Results: MODEL 1
gen school = (at16==1)
logit school able7 loginc ctratio oldsib yngsib etot female

```
Logistic regression                       Number of obs     = 3,423
                                          LR chi2(7)        = 964.33
                                          Prob > chi2       = 0.0000
Log likelihood = -1645.6347               Pseudo R2         = 0.2266
```

school	Coef.	Std. Err.	z	P>z
able7	.0407739	.0031581	12.91	0.000
loginc	.6760199	.11748	5.75	0.000
ctratio	-.2428075	.0241752	-10.04	0.000
oldsib	-.2935737	.0743979	-3.95	0.000
yngsib	-.0634156	.0379222	-1.67	0.094
etot	.221943	.0170205	13.04	0.000
female	-.0590708	.0858011	-0.69	0.491
_cons	-3.795431	.6840002	-5.55	0.000

(a) Estimate probability to stay at school for the student with
able7=50, loginc=5, ctratio=10, oldsib=yngsib=0, etot=4, female=0.
(b) Estimate marginal effect of increasing *income* by 5% on probability to stay at school for the student (a).
(c) Estimate marginal effect of increasing *yngsib* by 1 on probability to stay at school for the student (a).

Results: MODEL 2
Results of ordered logit are:
ologit at16 able7 loginc ctratio oldsib yngsib etot female

```
Ordered logistic regression                     Number of obs     =      3,423
                                                LR chi2(7)        =    1059.69
                                                Prob > chi2       =     0.0000
Log likelihood = -3212.3484                     Pseudo R2         =     0.1416

--------------------------------------------------------
        at16 |      Coef.   Std. Err.      z    P>|z|
-------------+------------------------------------------
       able7 |  -.0262445   .0019909   -13.18   0.000
      loginc |  -.6430468   .0929624    -6.92   0.000
     ctratio |   .1399681   .0181842     7.70   0.000
      oldsib |   .2731451   .0561998     4.86   0.000
      yngsib |   .0827978   .0283587     2.92   0.004
        etot |  -.1878267   .0121532   -15.45   0.000
      female |   .5831287   .0687306     8.48   0.000
-------------+------------------------------------------
       /cut1 |  -3.677982   .5132702
       /cut2 |  -2.059902   .5105434
--------------------------------------------------------
```

Use ordered logit to estimate:

(d) Estimate probability to stay at school for the student (a).

(e) Estimate probabilities of two other choices for that student.

(f) Estimate marginal effect of increasing *yngsib* by 1 on probability to stay at school for the student).

Solution

(a) $x'\hat{\beta} = 0.0407739 \cdot 50 + 0.6760199 \cdot 5 - 0.2428075 \cdot 10 + 0.221943 \cdot 4 - 3.795431 = 0.0831$

$p = \Lambda(0.0831) = 0.521$.

(b) $effect \approx \dfrac{\partial p}{\partial loginc} \ln(1.10) = \lambda(x'\hat{\beta}) \cdot 0.6760199 \cdot 0.04879 = 0.008232$.

(c) $x'_{(1)}\hat{\beta} = x'\hat{\beta} - 0.0634156 = 0.019645$. $\Delta p = \Lambda(0.019645) - \Lambda(0.0831) = 0.504911 - 0.520753 = -0.0158$.

(d) $P(at16 = 1) = \Lambda(c_1 - x'\hat{\beta}) = \Lambda(-3.677982 + 3.87908) = \Lambda(-0.2011) = 0.550$.

(e) $P(at16 = 2) = \Lambda(c_2 - x'\hat{\beta}) - \Lambda(c_1 - x'\hat{\beta}) = \Lambda(1.8192) - \Lambda(-0.2011) = 0.8605 - 0.550 = 0.310$.

$P(at16 = 3) = 0.140$.

(f) $\Delta p = \Lambda(c_1 - x'\hat{\beta} - 0.0827978) - \Lambda(c_1 - x'\hat{\beta}) = \Lambda(0.1183) - \Lambda(0.2011) = 0.5295 - 0.5501 = -0.02057$.

Задача 2

Crime1.xls, Wooldrige.

The data set contains data on arrests during 1986 on 2,725 men born in California in 1960 or 1961. Each man in the sample was arrested at least once prior to 1986. We are interested in explaining what determines how often these men are arrested again in 1986. The following variables are available.

narr86	# times arrested, 1986
avgsen	average sentence length served for prior convictions (in months)
black	=1 if black
born60	=1 if born in 1960
durat	recent unemployment duration (in months)
hispan	=1 if Hispanic
inc86	legal income, 1986, $100s
pcnv	proportion of prior convictions
ptime86	months in prison during 1986
qemp86#	quarters employed, 1986
tottime	time in prison since 18 (in months)

The variable *pcnv*, the proportion of arrests prior to 1986 that led to conviction, is a proxy for the likelihood of being convicted for a crime.

The variable *avgsen* is a proxy for the severity of punishment, if convicted.

The variable *ptime86* captures the incarcerative effect of crime: if an individual is in prison, he cannot be arrested for a crime outside of prison. Labour market opportunities are captured by *qemp86*.

Summary statistics:

	avgsen	black	born60	durat	hispan	inc86	pcnv	ptime86	qemp86	tottime
Mean	0.632	0.161	0.363	2.251	0.218	54.97	0.358	0.387	2.309	0.839
Median	0	0	0	0	0	29	0.25	0	3	0
Maximum	59.2	1	1	25	1	541	1	12	4	63.4
Minimum	0	0	0	0	0	0	0	0	0	0
Std. Dev.	3.508	0.368	0.481	4.607	0.413	66.63	0.395	1.950	1.610	4.607

Tabulation of narr86

Value	Count	Percent	Cumulative Count	Cumulative Percent
0	1970	72.29	1970	72.29
1	559	20.51	2529	92.81
2	121	4.44	2650	97.25
3	42	1.54	2692	98.79
4	12	0.44	2704	99.23
5	13	0.48	2717	99.71
6	4	0.15	2721	99.85
7	1	0.04	2722	99.89
9	1	0.04	2723	99.93
10	1	0.04	2724	99.96
12	1	0.04	2725	100.00
Total	2725	100.00	2725	100.00

Probit-model for the probability at least one arrest was estimated:

Dependent Variable: NARR86>0				**MODEL 1**
Method: ML - Binary Probit (Quadratic hill climbing)				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	–0.323473	0.057875	–5.589133	0.0000
PCNV	–0.553008	0.071956	–7.685374	0.0000
AVGSEN	0.003730	0.007684	0.485502	0.6273
PTIME86	–0.081029	0.018028	–4.494518	0.0000
QEMP86	0.010286	0.024660	0.417109	0.6766
INC86	–0.004832	0.000686	–7.045535	0.0000
BLACK	0.467772	0.072006	6.496325	0.0000
HISPAN	0.288913	0.065449	4.414352	0.0000
Mean dependent var	0.277064	S.D. dependent var		0.447631
S.E. of regression	0.428541	Akaike info criterion		1.094808
Sum squared resid	498.9705	Schwarz criterion		1.112160
Log likelihood	–1483.677	Hannan–Quinn criter.		1.101080
Restr. log likelihood	–1608.184	Avg. log likelihood		–0.544468
LR statistic (7 df)	249.0143	McFadden R–squared		0.077421
Probability(LR stat)	0.000000			
Obs with Dep=0	1970	Total obs		2725
Obs with Dep=1	755			

Coefficient covariance matrix

	C	PCNV	AVGSEN	PTIME86	QEMP86	INC86	BLACK	HISPAN
C	0.003350	-0.001665	-1.97E-05	-0.000160	-0.000781	3.38E-06	-0.001538	-0.001059
PCNV	-0.001665	0.005178	-1.86E-05	-5.04E-05	-2.61E-05	3.07E-06	0.000188	-0.000140
AVGSEN	-1.97E-05	-1.86E-05	5.90E-05	-2.79E-05	-1.08E-06	2.21E-07	-5.15E-05	-1.52E-05
PTIME86	-0.000160	-5.04E-05	-2.79E-05	0.000325	7.34E-05	-5.52E-07	-6.59E-05	-7.16E-05
QEMP86	-0.000781	-2.61E-05	-1.08E-06	7.34E-05	0.000608	-1.21E-05	8.12E-05	-7.44E-05
INC86	3.38E-06	3.07E-06	2.21E-07	-5.52E-07	-1.21E-05	4.70E-07	2.30E-06	1.49E-06
BLACK	-0.001538	0.000188	-5.15E-05	-6.59E-05	8.12E-05	2.30E-06	0.005185	0.001242
HISPAN	-0.001059	-0.000140	-1.52E-05	-7.16E-05	-7.44E-05	1.49E-06	0.001242	0.004284

(a) Use Model 1 to estimate probabilities to be arrested, *P1*, in 1986 for the two men, with values of the variables: avgsen = 0, inc86 = 50, pcnv = 0.1, ptime86 = 0, qemp86 = 3, one of which is white and another is hispanic.

(b) find 95% confidence interval for the probability of Hispanic man from (a) to be arrested.

(c) Estimate the marginal effect of income, $\dfrac{\partial P0}{\partial\ inc86}$ for the hispanic man in (a). $P0 = 1 - P1$ — probability of no arrests.

Solution

$P1 = P(NARR86 > 0) = \Phi(x'\beta)$.

For white

$x'\beta = -0.323473\times 1 - 0.553008\times 0.1 + 0.00373\times 0 - 0.081029\times 0 + 0.010286\times 3 - 0.004832\times 50 +$

$+0.467772\times 0 + 0.288913\times 0 = -0.58952$

$P1 = \Phi(-0.58952) = 0.2777 = 27.8\%$.

For Hispanic in a similar way

$P1 = \Phi(-0.30060) = 0.3819 = 38.2\%$.

(b) $V(P1) = [\varphi(-0.30060)]^2 \cdot x'Vx = [0.381319]^2 \cdot 0.0035934$ = 0.00052. (use Excel or STATA)

(here $x' = [1\ 0.1\ 0\ 0\ 3\ 50\ 0\ 1]$ and V is given)

Confidence interval is

$P1 = 0.2778 \pm \sqrt{0.00052}\cdot 1.96 = [0.233,\ 0.323] = [23.3\%,\ 32.3\%]$.

(c) $\dfrac{\partial P0}{\partial\ inc86} = \dfrac{\partial(1-\Phi(x'\beta))}{\partial\ inc86} = -\varphi(x'\beta)\beta_{inc86} = -\varphi(-0.30060)\cdot(-0.004832) = 0.00184$.

Задача 3

Докажите, что для logit- модели функция правдоподобия имеет единственный максимум.

Solution

$$\ln L = \sum (y_i \ln\Lambda(x_i\beta) + (1-y_i)\ln(1-\Lambda(x_i\beta));$$

$$\frac{\partial}{\partial\beta'}\ln L = \sum\left(y_i \frac{\lambda(x_i'\beta)}{\Lambda(x_i'\beta)} - (1-y_i)\frac{\lambda(x_i'\beta)}{1-\Lambda(x_i'\beta)} \ln(1-\Lambda(x_i'\beta)\right) x_i' =$$

$$= \sum\left(\frac{\lambda(x_i'\beta)}{\Lambda(x_i'\beta)1-\Lambda(x_i'\beta)}(y_i(1-\Lambda(x_i'\beta)) - (1-y_i)\Lambda(x_i'\beta))\right)x_i' = \sum(y_i - \Lambda(x_i'\beta))x_i';$$

$$\frac{\partial^2}{\partial\beta\partial\beta'}\ln L = \frac{\partial}{\partial\beta}\sum(y_i - \Lambda(x_i'\beta))x_i' = -\sum\lambda(x_i'\beta)x_i x_i'$$ — negative semidefinite matrix. QED.