ДЗ-04 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

(a) Use logit, ordered logit, multinomial logit models to estimate probability *p* of choice 3 (regular job).
(b) Compare precisions of the 3 model prediction.
(c) Compare the probability of choice 3 for boys and girls with same characteristic.
(d) Calculate *dp* — change in *p* (see 3) if income increase by 50%. Plot *dp* against *p* in all 3 models.
(e) Comment your findings.

Problem 2

The binary choice model is $P(y_t=1)=F(\alpha+\beta d_t)$, where d_t — is dummy variable. 100 observations on $(d_t,\ y_t)$ are presented at the table.

		y	
		0	1
d	0	10	30
	1	40	20

(a) Estimate coefficients α, β, with *logit*-model.
(b) Test the hypothesis **H$_0$**: $\beta = 0$. Find P-value. (*Hint*: Use likelihood ratio test.)
(c) Estimate covariance matrix of the random vector $(\hat\alpha,\ \hat\beta)'$. Use it to test hypothesis in (b). Find P-value.

Solution

(a) Likelihood function is $L(\alpha,\beta)=F(\alpha)^{30}F(\alpha+\beta)^{20}(1-F(\alpha))^{10}(1-F(\alpha+\beta))^{40}$. Let $\gamma=\alpha+\beta$.

Since ML estimators are invariant we can estimate α,γ.

Log-likelihood is $\ln L(\alpha,\gamma)=30\ln F(\alpha)+20\ln F(\gamma)+10\ln(1-F(\alpha))+40\ln(1-F(\gamma))$.

F.O.C. are (here pdf $f(x)=F'(x)$.

$$\begin{cases} \dfrac{\partial \ln L}{\partial \alpha} = 30\dfrac{f(\alpha)}{F(\alpha)} - 10\dfrac{f(\alpha)}{1-F(\alpha)} = 0, \\ \dfrac{\partial \ln L}{\partial \gamma} = 20\dfrac{f(\gamma)}{F(\gamma)} - 40\dfrac{f(\gamma)}{1-F(\gamma)} = 0. \end{cases}$$

Solution is $F(\alpha) = 3/4, \quad F(\gamma) = 1/3$.

Since $F = F(z) = \dfrac{e^z}{1+e^z} = \dfrac{1}{1+e^{-z}}$, we have $1 + e^{-z} = \dfrac{1}{F}, \quad e^{-z} = \dfrac{1}{F} - 1, \quad z = -\ln\left(\dfrac{1}{F} - 1\right)$. Hence,

$\hat{\alpha} = -\ln\left(\dfrac{4}{3} - 1\right) = -\ln\left(\dfrac{1}{3}\right) = 1.0986$, $\hat{\gamma} = -\ln(3-1) = -\ln(2) = -0.6932$, $\hat{\beta} = \hat{\gamma} - \hat{\alpha} = -1.7918$.

(b) For the restricted model $\beta = 0$, we have $\ln L(\alpha) = 50\ln F(\alpha) + 50\ln(1 - F(\alpha))$,

$\dfrac{d\ln L(\alpha)}{d\alpha} = 50\dfrac{f(\alpha)}{F(\alpha)} - 50\dfrac{f(\alpha)}{1-F(\alpha)} = 0$, Solution is $F(\alpha) = \dfrac{1}{2}$, and $\ln L_R = 50\ln\dfrac{1}{2} + 50\ln\dfrac{1}{2} = -69.3147$.

Unrestricted log likelihood is $\ln L_{UR} = 30\ln\dfrac{3}{4} + 20\ln\dfrac{1}{3} + 10\ln\dfrac{1}{4} + 40\ln\dfrac{2}{3} = -60.6843$.

$LR = 2(\ln L_{UR} - \ln L_R) = 17.26 > \chi^2_{0.05}(1) = 3.84$, Hence **$H_0$**: $\beta = 0$ is rejected. P-value is $3.26 \cdot 10^{-5}$.

(c) $\ln f(y, x, \beta) = y\ln F(x'\beta) + (1-y)\ln(1 - F(x'\beta))$,

$\dfrac{\partial \ln f(y,x,\beta)}{\partial \beta} = \dfrac{y - F(x'\beta)}{F(x'\beta)(1-F(x'\beta))} f(x'\beta)x = (y - F(x'\beta))x$, (only for logit model)

$V_{as}^{-1} = J(\beta) = E\{(y - F(x'\beta))^2 xx'\}$.

$$\hat{V}_{as}^{-1} = \frac{1}{100}\left\{10(-F(\hat{\alpha}))^2\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 40(-F(\hat{\gamma}))^2\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix} + 30(1-F(\hat{\alpha}))^2\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 20(1-F(\hat{\gamma}))^2\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\right\} =$$

$$= \frac{1}{100}\left\{\left(10\frac{3^2}{4^2} + 30\frac{1^2}{4^2}\right)\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + \left(40\frac{1^2}{3^2} + 20\frac{2^2}{3^2}\right)\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\right\} = \frac{1}{100}\left\{7.5\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 13.33\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\right\} = \frac{1}{100}\begin{bmatrix}20.83 & 13.33\\13.33 & 13.33\end{bmatrix}.$$

$$\hat{V}\left(\begin{bmatrix}\hat{\alpha}\\\hat{\beta}\end{bmatrix}\right) = \frac{1}{100}\hat{V}_{as} = \frac{1}{100}\left\{\frac{1}{100}\begin{bmatrix}20.83 & 13.33\\13.33 & 13.33\end{bmatrix}\right\}^{-1} = \frac{1}{100}100\begin{bmatrix}20.83 & 13.33\\13.33 & 13.33\end{bmatrix}^{-1} = \begin{bmatrix}0.133 & -0.133\\-0.133 & 0.2083\end{bmatrix}.$$

$t_{\hat{\beta}} = \dfrac{\hat{\beta}}{s_{\hat{\beta}}} = -\dfrac{1.7918}{\sqrt{0.2083}} = -3.926$. Hence **$H_0$**: $\beta = 0$ is rejected. P-value is $2P(Z > 3.926) = 8.65 \cdot 10^{-5}$

Problem 3

Below you can find estimation results from a multinomial logit model. The dependent variable is
status = 0, if if enrolled in school; (College)
= 1, if not in school and not working (Home);
= 2, if working (Work).

Choice 0 (College) is the base choice in the model. Regressors are number of years of education at the current moment (*educ*), experience of employment (*exper*), and its square (*exper*^{2}), also a race dummy *black* = 1, if the person is afro American.

Table 1: Multinomial Logit Estimates of School and Labor Market Decisions

Explanatory Variable	Home (status=1)	Work (status=2)
educ	-0.674	-0.315
	(0.070)	(0.065)
exper	-0.106	0.849
	(0.173)	(0.157)
exper^2	-0.013	-0.077
	(0.025)	(0.023)
black	0.813	0.311
	(0.303)	(0.282)
cons	10.28	5.54
	(1.13)	(1.09)
# of obs	1717	
% correctly predicted	79.6	
Log-likelihood value	-907.86	

Pseudo- R^2	0.243

(a) (5 points) Give interpretation of the coefficients at *educ* in both columns of the table (Home, Work). How *educ* influence on the probability to be in school?

(b) (5 points) Calculate the marginal effect $\dfrac{\partial P(y=0\,|\,x)}{\partial\, educ}$ for a white person with 12 years of schooling and 10 years of experience.

(c) (5 points) Describe the model for which the property "Independence of irrelevant alternatives" is relevant. What is "Independence of irrelevant alternatives"?

Solution

(a) *Home*: With additional year of education, odds ratio $P(y=1\,|\,x)/P(y=0\,|\,x)$ decreases since the coefficient at *educ* is negative (–0.674). *Work*: odds ratio $P(y=2\,|\,x)/P(y=0\,|\,x)$ also decreases since the coefficient at *educ* is negative (–0.315). So far as $P(y=2\,|\,x)+P(y=1\,|\,x)+P(y=0\,|\,x)\equiv 1$, we have

$\dfrac{P(y=2\,|\,x)}{P(y=0\,|\,x)}+\dfrac{P(y=1\,|\,x)}{P(y=0\,|\,x)}+1\equiv\dfrac{1}{P(y=0\,|\,x)}$, hence, the probability of choice 0, $P(y=0\,|\,x)$ increases.

(b) $\dfrac{\partial P(y=0\,|\,x)}{\partial\, educ}=\dfrac{\partial}{\partial x_1}\dfrac{1}{1+\sum_{k=1}^{2}\exp(x'\beta_k)}=\dfrac{-\left(\exp(x'\beta_1)\beta_{11}+\exp(x'\beta_2)\beta_{21}\right)}{\left(1+\sum_{k=1}^{2}\exp(x'\beta_k)\right)^2}$; $\exp(x'\beta_1)=\exp(-0.168)=0.845$,

$\exp(x'\beta_2)=\exp(2.550)=12.807$. $\dfrac{\partial P(y=0\,|\,x)}{\partial\, educ}=-\dfrac{0.845\cdot(-0.674)+12.807\cdot(-0.315)}{(1+0.845+12.807)^2}=0.021$.

(c) Under the multinomial logit model, the probability of selecting choice i is $P_j=\dfrac{e^{V_i}}{\sum_{k\in C}e^{V_k}}$, were random utility $U_i=V_i+\varepsilon_i$. Thus $\dfrac{P_i}{P_j}=e^{V_i-V_j}$, the ratio of probabilities depends only on choices i and j, but not on any other alternatives. This follows from the assumption of i.i.d. ε_i.