PRINCIPLES OF ECONOMETRICS

5TH EDITION

ANSWERS TO ODD-NUMBERED EXERCISES IN CHAPTER 16

EXERCISE 16.1

(a) Using the logit estimates, the probability of a person choosing automobile transportation given that *DTIME* = 1 is $\Lambda(\tilde{\gamma}_1 + \tilde{\gamma}_2 DTIME) = 0.5729$

(b) Using the probit estimates, the probability of a person choosing automobile transportation given that *DTIME* = 1 is $\Phi(\tilde{\beta}_1 + \tilde{\beta}_2 DTIME) = 0.5931$. There is very little difference between the logit and probit predicted probabilities.

(c) Given that *DTIME* = 3, the marginal effect of an increase in *DTIME* using the logit estimates is

$$\widehat{\frac{dp}{dx}} = \lambda(\tilde{\gamma}_1 + \tilde{\gamma}_2 30)\tilde{\gamma}_2 = 0.086537$$

The estimated linear probability model is in Example 16.2. The marginal effect is the estimated slope of 0.0703, which is a bit smaller than the logit estimate. But the important point is that for the linear probability model the marginal effect is constant and doesn't depend on the existing commute time.

(d) The logit estimate is $\widehat{\frac{dp}{dx}} = \lambda(\tilde{\gamma}_1 + \tilde{\gamma}_2 \times -5)\tilde{\gamma}_2 = 0.02642$.

The probit estimate is $\widehat{\frac{dp}{dx}} = \phi(\tilde{\beta}_1 + \tilde{\beta}_2 \times -5)\tilde{\beta}_2 = 0.035204$

EXERCISE 16.3

(a) $P(y = 1 \mid x = 1.5) = \Phi(\tilde{\beta}_1 + \tilde{\beta}_2 1.5) = 0.953996$

(b) Using the threshold 0.5 we predict $y = 1$ based on the results in part (a). This does agree with the data value.

(c) The likelihood function is

$$L(\beta_1, \beta_2 \mid \mathbf{y}, \mathbf{x}) = \Phi(-0.70)\Phi(-0.88)\left[1 - \Phi(-0.86)\right] = 0.0369$$

The corresponding log likelihood is

$$\ln L(\beta_1, \beta_2 \mid \mathbf{y}, \mathbf{x}) = \ln(0.0369) = -3.2995$$

This is smaller than the log-likelihood evaluated at the maximum likelihood estimates.

(d) True. Each term is in the likelihood function is a probability, $\Phi(\text{something})$. This term will always be between zero and one, so that the product of such terms will also between zero and one.

(e) True. The logarithm of any number less than one is negative.

EXERCISE 16.5

(a) $$L(\gamma_1 \mid \mathbf{y},\mathbf{x}) = \Lambda(\gamma_1)^{y_1}\left[1-\Lambda(\gamma_1)\right]^{1-y_1}\Lambda(\gamma_1)^{y_2}\left[1-\Lambda(\gamma_1)\right]^{1-y_2}\Lambda(\gamma_1)^{y_3}\left[1-\Lambda(\gamma_1)\right]^{1-y_3}$$
$$= \Lambda(\gamma_1)^{1}\left[1-\Lambda(\gamma_1)\right]^{0}\Lambda(\gamma_1)^{1}\left[1-\Lambda(\gamma_1)\right]^{0}\Lambda(\gamma_1)^{0}\left[1-\Lambda(\gamma_1)\right]^{1}$$
$$= \Lambda(\gamma_1)\Lambda(\gamma_1)\left[1-\Lambda(\gamma_1)\right]$$
$$= \left[\Lambda(\gamma_1)\right]^2\left[1-\Lambda(\gamma_1)\right]$$

Then,

$$\ln L(\gamma_1 \mid \mathbf{y},\mathbf{x}) = \ln\left\{\left[\Lambda(\gamma_1)\right]^2\left[1-\Lambda(\gamma_1)\right]\right\} = 2\ln\Lambda(\gamma_1) + \ln\left[1-\Lambda(\gamma_1)\right]$$

(b) $$\frac{d\ln L(\gamma_1 \mid \mathbf{y},\mathbf{x})}{d\gamma_1} = 2\frac{d\ln\Lambda(\gamma_1)}{d\gamma_1} + \frac{d\ln\left[1-\Lambda(\gamma_1)\right]}{d\gamma_1}$$
$$= 2\frac{\Lambda'(\gamma_1)}{\Lambda(\gamma_1)} + \frac{-\Lambda'(\gamma_1)}{1-\Lambda(\gamma_1)} = 2\frac{\lambda(\gamma_1)}{\Lambda(\gamma_1)} + \frac{-\lambda(\gamma_1)}{1-\Lambda(\gamma_1)}$$

(c) Maybe. The second derivative must be negative at the value obtained from solving $d\ln L(\gamma_1)/d\gamma_1 = 0$.

(d) True. If $d\ln L(\gamma_1)/d\gamma_1 = 0$ at the value $\tilde{\gamma}_1$ and the second derivative is always negative, then $\tilde{\gamma}_1$ is the maximum of the log-likelihood function and the maximum likelihood estimate.

(e) $$\frac{d\ln L(\gamma_1 \mid \mathbf{y},\mathbf{x})}{d\gamma_1} = 2\frac{\lambda(\gamma_1)}{\Lambda(\gamma_1)} + \frac{-\lambda(\gamma_1)}{1-\Lambda(\gamma_1)} = 0$$
$$\Rightarrow \quad 2\frac{\lambda(\tilde{\gamma}_1)\left[1-\Lambda(\tilde{\gamma}_1)\right]}{\Lambda(\tilde{\gamma}_1)\left[1-\Lambda(\tilde{\gamma}_1)\right]} + \frac{-\lambda(\tilde{\gamma}_1)\Lambda(\tilde{\gamma}_1)}{\Lambda(\tilde{\gamma}_1)\left[1-\Lambda(\tilde{\gamma}_1)\right]} = 0$$
$$\Rightarrow \quad 2\lambda(\tilde{\gamma}_1)\left[1-\Lambda(\tilde{\gamma}_1)\right] - \lambda(\tilde{\gamma}_1)\Lambda(\tilde{\gamma}_1) = \lambda(\tilde{\gamma}_1)\left\{2\left[1-\Lambda(\tilde{\gamma}_1)\right] - \Lambda(\tilde{\gamma}_1)\right\} = 0$$
$$\Rightarrow \quad 2\left[1-\Lambda(\tilde{\gamma}_1)\right] - \Lambda(\tilde{\gamma}_1) = 2 - 3\Lambda(\tilde{\gamma}_1) = 0$$
$$\Rightarrow \quad \Lambda(\tilde{\gamma}_1) = 2/3$$

(f) As just shown, the condition in (d) implies $\Lambda(\tilde{\gamma}_1) = 2/3$. Then,

$$\frac{1}{1+e^{-\tilde{\gamma}_1}} = \frac{2}{3} \quad \Rightarrow \quad 2\left[1+e^{-\tilde{\gamma}_1}\right] = 3 \quad \Rightarrow \quad 1+e^{-\tilde{\gamma}_1} = 3/2 \quad \Rightarrow \quad e^{-\tilde{\gamma}_1} = 1/2$$
$$\Rightarrow \quad -\tilde{\gamma}_1 = \ln(1/2) \quad \Rightarrow \quad \tilde{\gamma}_1 = -\ln(1/2)$$

EXERCISE 16.7

(a) Using $\Lambda\left(\tilde{\gamma}_1\right) = 2/3$ we have

$$\ln L\left(\tilde{\gamma}_1\right) = 2\ln\left(2/3\right) + \ln\left[1 - 2/3\right] = 2\left(-0.40546511\right) - 1.0986123 = -1.9095425$$

Using $\tilde{\gamma}_1 = -\ln\left(1/2\right) = 0.6931472$ we have

$$\begin{aligned}\ln L\left(\tilde{\gamma}_1\right) &= 2\ln\Lambda\left(\tilde{\gamma}_1\right) + \ln\left[1 - \Lambda\left(\tilde{\gamma}_1\right)\right] = 2\ln\left(0.66666667\right) + \ln\left(0.33333333\right) \\ &= 2\left(-0.4054651\right) - 1.0986123 = -1.9095425\end{aligned}$$

(b) The likelihood ratio statistic is

$$LR = 2\left(\ln L_u - \ln L_R\right) = 2\left(-1.612 - (-1.9095425)\right) = 0.59508501$$

The test critical value is $\chi^2_{(0.95,1)} = 3.8415$; we fail to reject the null hypothesis that $H_0 : \gamma_2 = 0$.

(c) The *p* value is $p = 1 - PROB\left[\chi^2_{(1)} \le 0.59508501\right] = 0.44045949$

EXERCISE 16.9

(a) We see that the larger *GRADES* (the worse the performance) the probability of attending college falls. The *t*-value for the coefficient is −11.5655 which is significant at the 5% level, using the critical values ±1.96. On the other hand, a higher family income increases the probability of attending college. The *t* = 4.4668 so that it is also statistically significant. The coefficient of *BLACK* is positive indicating that black students have a higher probability of attending college holding other factors equal. The *t* = 2.9472, which is also significant.

(b) The first probability is $P\left(COLLEGE = 1\right) = \Phi\left(2.3321\right) = 0.9902$

The second is $P\left(COLLEGE = 1\right) = \Phi\left(1.4117\right) = 0.9210$

(c) $P\left(COLLEGE = 1\right) = \Phi\left(2.0533\right) = 0.9800$. The estimated probability of attending college is about 0.06 higher for the black student.

(d) Evaluating the marginal effect at the given values we obtain

$$\begin{aligned}\frac{\partial P\left(COLLEGE = 1\right)}{\partial FAMINC} &= \phi\left(2.5757 - 0.3068\left(5\right) + 0.0074\left(50\right) + 0.6416\left(0\right)\right)0.0074 \\ &= 0.00109\end{aligned}$$

We estimate that the marginal effect of an increase in family income, given the values of the explanatory variables, increases the probability of attending college by 0.0011.

(e) The value of the likelihood ratio test statistic is $LR = 2\left(\ln L_U - \ln L_R\right) = 29.80$. The test critical value is $\chi^2_{(0.99,2)} = 9.2103$. Therefore, we reject the null hypothesis that *FAMINC* and *BLACK* have no effect.

EXERCISE 16.11

(a) In model 1 we estimate that as the mother ages the probability of a low birthweight baby declines. The *t*-value is −1.90, which is significant at the 10% level, but not at 5%. Having a prenatal visit in the first trimester reduces the probability of a low birthweight baby. For this coefficient $t = -2.21$, so that it is significant at 5%. Mothers who smoke have a higher probability of a low birthweight baby. The $t = 5.93$ on this coefficient; it is significant at better than the 0.01 level. In fact, the two-tail *p*-value is 3.029×10^{-9}.

(b) The marginal effects are

$$\widehat{ME}_{20} = \phi\left(-1.2581 - 0.0103(20) - 0.1568(0) + 0.3974(0)\right)(-0.0103) = -0.00140693$$

$$\widehat{ME}_{50} = \phi\left(-1.2581 - 0.0103(50) - 0.1568(0) + 0.3974(0)\right)(-0.0103) = -0.00085322$$

These results suggest that, for both women, the marginal effect of a year of age decreases the probability of a low birth weight baby, with the decrease greater for the younger woman. One might suspect an increase in the probability for the older woman, in which case the results do not make sense.

(c) The marginal effects are

$$\begin{aligned}\widehat{ME}_{20} &= \phi\left(-0.1209 - 0.1012(20) - 0.1387(0) + 0.4061(0) + 0.0017\left(20^2\right)\right)\times \\ &\quad \left(-0.1012 + 2(0.0017)(20)\right) \\ &= -0.00452965\end{aligned}$$

$$\begin{aligned}\widehat{ME}_{50} &= \phi\left(-0.1209 - 0.1012(20) - 0.1387(0) + 0.4061(0) + 0.0017\left(50^2\right)\right)\times \\ &\quad \left(-0.1012 + 2(0.0017)(50)\right) \\ &= 0.01779607\end{aligned}$$

Now we find that the marginal effect of a year of age for a 20 year old is negative, but the marginal effect for a 50 year old is to increase the probability of a low birthweight baby. This is more reasonable compared to the results in part (b).

(d) The difference in probabilities is

$$\begin{aligned}\Delta P(LBWEIGHT) &= \Phi\left(-0.1209 - 0.1012(30) - 0.1387(1) + 0.4061(1) + 0.0017\left(30^2\right)\right) \\ &\quad - \Phi\left(-0.1209 - 0.1012(30) - 0.1387(0) + 0.4061(1) + 0.0017\left(30^2\right)\right) \\ &= -0.02408678\end{aligned}$$

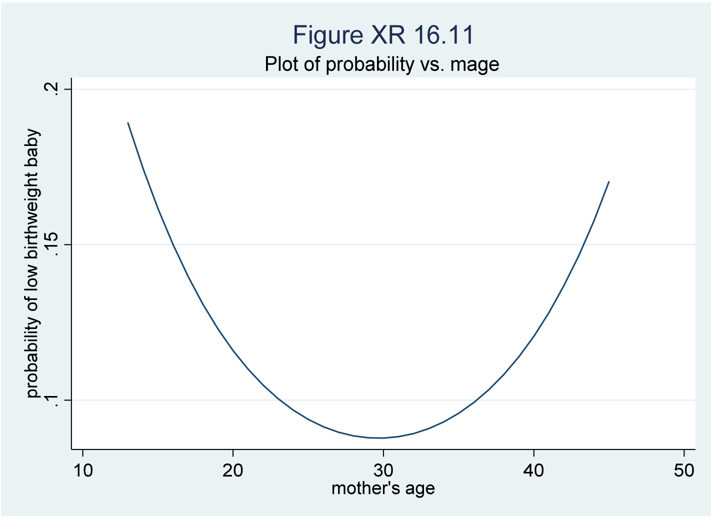
(e) The difference in the probabilities is

$$\begin{aligned}\Delta P(LBWEIGHT) &= \Phi\left(-0.1209 - 0.1012(30) - 0.1387(1) + 0.4061(1) + 0.0017\left(30^2\right)\right) \\ &\quad - \Phi\left(-0.1209 - 0.1012(30) - 0.1387(1) + 0.4061(0) + 0.0017\left(30^2\right)\right) \\ &= 0.04826261\end{aligned}$$

(f) The answer comes from setting the following derivative equal to zero and solving for *MAGE*.

$$\frac{\partial P(LBWEIGHT)}{\partial MAGE} = \phi\left(\beta_1 + \beta_2 MAGE + \beta_3 PRENATAL1 + \beta_4 MBSMOKE + \beta_5 MAGE^2\right) \times (\beta_2 + 2\beta_5 MAGE) = 0$$

The first component $\phi\left(\beta_1 + \beta_2 MAGE + \beta_3 PRENATAL1 + \beta_4 MBSMOKE + \beta_5 MAGE^2\right)$ is never zero. It is the standard normal *pdf*. For the derivative to be zero it must be the case that $(\beta_2 + 2\beta_5 MAGE) = 0$, or that $MAGE = -\beta_2 / 2\beta_5$. Solving for this value using the estimates yields $MAGE^* = 29.7647$, or just slightly less than 30 years of age. For this to minimize the probability the second derivative must be positive. That is a bit tedious, so instead we plot the probability. The plot is convex, so the value we found is the minimizing value.



EXERCISE 16.13

(a) The conditional logit probability is given in equation (16.23) with the modification that the numerator is

$$\exp\left(\beta_2 PRICE_{ij} + \beta_3 FEATURE_{ij} + \beta_4 DISPLAY_{ij}\right)$$

The odds of Coke relative to Pepsi at the given prices are

$$\frac{\exp(\beta_2 PRICE_{i3} + \beta_4)}{\exp(\beta_2 PRICE_{i1})} = \exp\left(\beta_2 \left[PRICE_{i3} - PRICE_{i1}\right] + \beta_4\right)$$

Inserting the estimates, we have

$$\exp\left(-1.7445[1.25 - 1.25] + 0.4624\right) = \exp(0.4624) = 1.58788$$

Under this scenario Coke is about 1.5 time as likely to be chosen as Pepsi.

(b) Under this scenario the odds are very close to even.

$$\exp\left(-1.7445[1.25 - 1.00] + 0.4624\right) = 1.0266$$

(c) The numerator of each probability is

$$\exp\left(\beta_2 PRICE_{ij} + \beta_3 FEATURE_{ij} + \beta_4 DISPLAY_{ij}\right)$$

Calculating these terms for the two Coke prices and for Pepsi and 7-Up.

Coke @ \$1.25: $\exp\left(-1.7445(1.25)+0.4624\right)=0.17938427$

Coke @ \$1.50: $\exp\left(-1.7445(1.50)+0.4624\right)=0.11597855$

Pepsi @ \$1.25: $\exp\left(-1.7445(1.25)\right)=.1129709$

7-Up @ \$1.25: $\exp\left(-1.7445(1.25)\right)=.1129709$

Thus, the Coke probability when the price is \$1.25 is 0.4426 and when it is \$1.50 it falls to 0.3392, a change of 0.1034.

(d) The odds ratio is

$$\frac{\exp\left(\beta_2 PRICE_{i3} + \beta_4\right)}{\exp\left(\beta_{12} + \beta_2 PRICE_{i1}\right)} = \exp\left(\beta_2\left[PRICE_{i3} - PRICE_{i1}\right] + \beta_4 - \beta_{12}\right)$$

Inserting the estimates, we have

$$\exp\left(-1.8492\left[PRICE_{i3} - PRICE_{i1}\right] + 0.4727 - 0.2841\right) = 1.2076$$

(e) Following the procedure in part (c), but adding in the alternative specific constants for Pepsi and 7-Up, we calculate these terms for the two Coke prices and for Pepsi and 7-Up.

Coke @ \$1.25: $\exp\left(-1.8492(1.25)+0.4624\right)=0.15737874$

Coke @ \$1.50: $\exp\left(-1.8492(1.50)+0.4624\right)=0.09912238$

Pepsi @ \$1.25: $\exp\left(-1.84925(1.25)+0.2841\right)=0.13167744$

7-Up @ \$1.25: $\exp\left(-1.8492(1.25)+0.0907\right)=0.10851547$

Thus, the Coke probability when the price is \$1.25 is 0.3959 and when it is \$1.50 it falls to 0.2921, a change of 0.1038.

(f) The likelihood ratio statistic is

$$LR = 2\left(\ln L_U - \ln L_R\right) = 2\left(-1811.3543 - \left(-1824.5621\right)\right) = 26.4156$$

The test critical value is $\chi^2_{(0.99,2)} = 9.2103$. Thus, we reject the null hypothesis that both coefficients are zero.

EXERCISE 16.15

(a) Using the hint, we estimate that a 1% increase in GDP will increase the number of medals won by 0.5556%. To show that the coefficient is an elasticity consider the following simpler model,

$$E(y|\mathbf{x}) = \exp(\beta_1 + \beta_2 \ln(x)) = x^{\beta_2} \exp(\beta_1)$$
$$\Rightarrow \quad \ln[E(y|\mathbf{x})] = \beta_1 + \beta_2 \ln(x)$$

In this way β_2 is the coefficient in a log-log model, which is a constant elasticity model.

(b) Their expected number of medals won was

$$E(y|GDP = 8.356, POP = 11.8) = \exp(\tilde{\beta}_1 + \tilde{\beta}_2 \ln(POP) + \tilde{\beta}_3 \ln(GDP)) = 1.47$$

(c) $\hat{P}(y \le 1) = \hat{P}(y = 0) + \hat{P}(y = 1) = 0.2299 + 0.3380 = 0.5679$

(d) $E(y|GDP = 306, POP = 6.875) = \exp(\tilde{\beta}_1 + \tilde{\beta}_2 \ln(POP) + \tilde{\beta}_3 \ln(GDP)) = 8.58$

(e) $\hat{P}(y \le 1) = \hat{P}(y = 0) + \hat{P}(y = 1) = 0.0002 + 0.0016 = 0.0018$

(f) Using the equation at the top of page 715,

$$100[e^{\delta} - 1] = 100[e^{0.6620} - 1] = 93.867$$

That is, being a host country is estimated to increase the expected number of medals won by 94%. That is huge! The coefficient of *HOST* has a $t = 4.81$ so it is statistically significant at the 1% level.

(g) Using Model 1

$$E(y|GDP = 7280, POP = 265) = \exp(\tilde{\beta}_1 + \tilde{\beta}_2 \ln(POP) + \tilde{\beta}_3 \ln(GDP)) = 109.11$$

Using Model 2

$$E(y|GDP = 7280, POP = 265) = \exp(\tilde{\beta}_1 + \tilde{\beta}_2 \ln(POP) + \tilde{\beta}_3 \ln(GDP) + \tilde{\beta}_4) = 208.08$$

In this case Model 1 did the better job of predicting the outcome.

EXERCISE 16.17

(a) These estimates are in Table XR16.17a, column (2). We see that none of the variables are significant, except the constant term. The likelihood ratio statistic is $LR = 2(\ln L_U - \ln L_R) = 2(-736.585 - (-737.224)) = 1.278$. The test critical value is $\chi^2_{(0.95,2)} = 5.9915$. Therefore, we fail to reject the null hypothesis that the variables *BOY* or *BLACK* has no effect on the probability of assignment to a small class. We cannot reject the null hypothesis that the assignment to small classes was done randomly.

(b) These results are in Tables XR16.17b and XR16.17c, column (2). Again, we find neither of the variables individually significant and the likelihood ratio test statistics for the two models are 0.58 and 0.15, indicating a lack of joint significance as well. We cannot reject the null hypothesis that the assignment to classes with an aide or a regular class was done randomly.

(c) In the Tables of results, the model including *FREELUNCH* is in column (3). Note that the coefficient of *FREELUNCH* is not significant in any of the models. The likelihood ratio test statistics of the models versus the model in column (1) containing a constant only are, for small classes 1.41, or classes with an aide 0.69 and for regular sized classes 0.15. The test critical value is $\chi^2_{(0.95,3)} = 7.8147$ so that we again fail to reject the joint null hypothesis. We cannot reject the null hypothesis that the assignment to small classes, or a class with an aide, or a regular sized class was done randomly.

Table XR16.17a logit models: *SMALL*

	(1) CONST	(2) LIST 1	(3) LIST 2	(4) LIST 3
C	-0.8275***	-0.7515***	-0.7347***	-0.9296***
	(-13.188)	(-7.600)	(-6.736)	(-4.198)
BOY		-0.0662	-0.0670	-0.0686
		(-0.527)	(-0.533)	(-0.543)
BLACK		-0.1375	-0.1130	-0.0453
		(-1.005)	(-0.741)	(-0.276)
FREELUNCH			-0.0512	-0.0707
			(-0.365)	(-0.501)
TCHWHITE				0.3632
				(1.811)
TCHMASTERS				-0.3677**
				(-2.694)
N	1200	1200	1200	1200
Log-likelihood	-737.224	-736.585	-736.519	-731.728
Chi-square	0.000	1.278	1.411	10.991

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table XR16.17b logit models: *AIDE*

	(1) CONST	(2) LIST 1	(3) LIST 2	(4) LIST 3
C	-0.6782***	-0.7336***	-0.7485***	-1.2237***
	(-11.102)	(-7.498)	(-6.928)	(-5.633)
BOY		0.0602	0.0609	0.0769
		(0.493)	(0.498)	(0.626)
BLACK		0.0773	0.0560	0.2058
		(0.589)	(0.382)	(1.305)
FREELUNCH			0.0445	0.0385
			(0.326)	(0.281)
TCHWHITE				0.4543*
				(2.328)
TCHMASTERS				0.0948
				(0.732)
N	1200	1200	1200	1200
Log-likelihood	-766.560	-766.270	-766.217	-762.833
Chi-square	0.000	0.580	0.686	7.453

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table XR16.17c logit models: *REGULAR*

	(1) CONST	(2) LIST 1	(3) LIST 2	(4) LIST 3
C	-0.5790***	-0.5958***	-0.5972***	-0.0133
	(-9.622)	(-6.215)	(-5.646)	(-0.066)
BOY		0.0024	0.0025	-0.0125
		(0.020)	(0.020)	(-0.103)
BLACK		0.0496	0.0477	-0.1641
		(0.383)	(0.330)	(-1.032)
FREELUNCH			0.0040	0.0281
			(0.030)	(0.207)
TCHWHITE				-0.7273***
				(-4.016)
TCHMASTERS				0.2431
				(1.890)
N	1200	1200	1200	1200
Log-likelihood	-783.525	-783.451	-783.451	-774.537
Chi-square	0.000	0.147	0.148	17.974

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(d) In the Tables of results, the model including *TCHWHITE* and *TCHMASTERS* is in column (4). Note that in the small class models *TCHWHITE* is significant at the 1% level. In the models for aide *TCHWHITE* is significant at the 5% level. And in the model for regular class assignment *TCHWHITE* is significant at the 0.1% level. Carrying out the likelihood ratio test of the model in column (4) versus the model in column (1) we find the LR statistics to be 10.99 (*SMALL*), 7.45 (*AIDE*) and 17.97 (*REGULAR*) and the critical value for these tests is $\chi^2_{(0.95,5)} = 11.0705$. Thus, we conclude that at least one of the variables in column (4) helps explain assignment. Carrying out the likelihood ratio test versus the models in column (3) we find the LR statistics to be 9.58 (*SMALL*), 6.77 (*AIDE*) and 17.83 (*REGULAR*). The test critical value is $\chi^2_{(0.95,2)} = 5.9915$. Thus for each classification we find evidence that *TCHWHITE* and/or *TCHMASTERS* helps explain the class-type assignment. Students were randomized within schools but not across schools. Some schools were in wealthier districts, or predominately white districts, so it is not surprising that these variables were significant.

EXERCISE 16.19

(a) The linear probability model results are in column (1) of Table XR16.19. The coefficients of the indicators for an adjustable rate mortgage or a refinance, *ARM* and *REFINANCE*, are not statistically significant. Having a second mortgage, *LIEN2*, has a positive, large and significant coefficient, suggesting that a second mortgage is a significant factor in predicting a loan default. We estimate that a second mortgage increases the probability of default by 0.2424. We cannot reject the null hypothesis that having a 30 year over a 15 year mortgage (the indicator *TERM30*) has no effect of the probability of default. The coefficient of the indicator variable *UNDERWATER* is positive and significant. We predict that a mortgage that is underwater increases the probability of default by 0.1896. A higher loan to value ratio, *LTV*, predicts a higher probability of default, as does a higher mortgage loan rate, *RATE*. The coefficient of *AMOUNT* is not statistically different from zero. A borrower's *FICO* score has a negative and significant effect. A 100 point larger *FICO* predicts the probability of default to be 0.14 lower. All the signs of the statistically significant coefficients are what we would expect.

Table XR16.19

	(1) OLS	(2) Probit
C	0.5664*	-0.1785
	(2.458)	(-0.254)
ARM	-0.0275	-0.0867
	(-0.851)	(-0.925)
REFINANCE	-0.0341	-0.1012
	(-1.109)	(-1.093)
LIEN2	0.2424**	0.9454**
	(3.087)	(3.220)
TERM30	0.0223	0.0712
	(0.508)	(0.534)
UNDERWATER	0.1896***	0.7925**
	(4.162)	(3.241)
LTV	0.0055***	0.0163***
	(4.379)	(3.833)
RATE	0.0443***	0.1269***
	(5.358)	(5.142)
AMOUNT	-0.0010	-0.0032
	(-1.101)	(-1.038)
FICO	-0.0014***	-0.0040***
	(-5.836)	(-5.293)
N	1000	1000

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) The probit estimates are in column (2) of Table XR16.19. All the variables that are significant have the same sign as in the LPM model, and most, other than the constant term and the coefficient of *UNDERWATER*, carry the same level of significance.

(c) The probabilities are given in the following table. Borrower 500 has a higher predicted probability of default in both models and in fact did default. The borrower was underwater and borrowed 80% of the value of the property. The loan was for about \$250,000 and the borrower did not have a particularly high *FICO* score. Borrower 1000 has a lower predicted probability of default in both models and in fact did not default. The borrower was underwater but borrowed only 63.83% of the value of the property. The loan was for about \$150,000 and the borrower did have a high *FICO* score. In this case the probit and linear probability model predictions of the probability of default were similar.

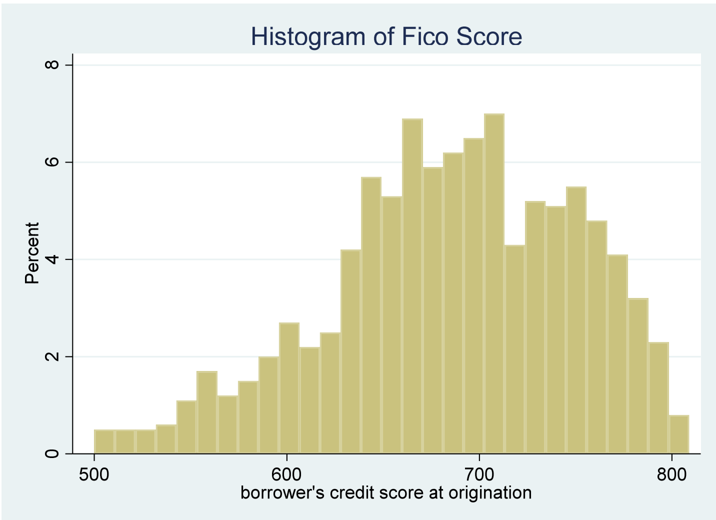
Observation	LPM	Probit
500	0.4299	0.4013
1000	0.2772	0.2625

(d) The histogram of the *FICO* scores for this set of borrower shows values between 500 and 800. The median value is 688.5, the 25th percentile is 647 and the 75th percentile 736.5. Thus, the values 500, 600 and 700 are very low, low (10th percentile is 596.5) and above average.

Given these loan characteristics, and using the linear probability model, the *FICO* score loan of 500 is predicted to have a 0.64 probability of default. The *FICO* score loan of 600 is predicted to have a 0.50 probability of default. The *FICO* score loan of 700 is predicted to have a 0.35 probability of default.

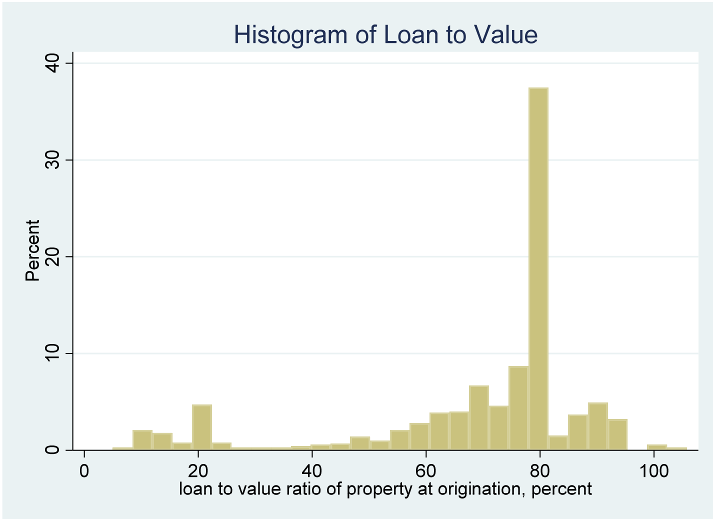
Given these loan characteristics, and using the probit model, the *FICO* score loan of 500 is predicted to have a 0.56 probability of default. The *FICO* score loan of 600 is predicted to have a 0.40 probability of default. The *FICO* score loan of 700 is predicted to have a 0.26 probability of default. The two models do not predict the same values.

For the linear probability model each 100 point increase in the *FICO* score reduces the predicted probability of default by 0.14. For the probit model the increase from 500 to 600 reduces the predicted probability of default by 0.16, while the increase from 600 to 700 reduces the predicted probability of default by 0.14.



(e) As noted in the previous solution, for the linear probability model *FICO* has a constant marginal effect, each point reducing the probability of default by 0.001436. For the probit model the marginal effect is not constant, as illustrated in Figure 16.1. For the probit model, the marginal effects are −0.0015671, −0.0015351 and −0.0012843 for *FICO* 500, 600 and 700 respectively. The higher the value of *FICO* the smaller the impact on the probability of additional increases.

(f) The histogram is given below. The most popular value is 80%, the mode, while 80% is also the 75th percentile of the *LTV* distribution. Using the linear probability model, the predicted default probabilities are 0.17 and 0.50 for the given scenarios, with *LTV* = 20 and 80 respectively. Using the probit model the corresponding predictions are 0.11 and 0.40, respectively. As anticipated, the larger loan to value indicates a riskier loan and results in higher predictions of default.



(g) Summaries of the predictions from the linear probability and probit models using a threshold of 0.5 are given in the following tables. The diagonal percentages are successes. The linear probability model predicted 53.4% of non-defaults correctly and 15.8% of the defaults. In the probit model the results are similar. In the sample of 1000 loans, 37.8% of the loans defaulted.

```
   = 1 if |
  payment |
  late by |      linear probability model
 90+ days |         0          1 |     Total
----------+----------------------+----------
        0 |       534         88 |       622
          |     53.40       8.80 |     62.20
----------+----------------------+----------
        1 |       220        158 |       378
          |     22.00      15.80 |     37.80
----------+----------------------+----------
    Total |       754        246 |     1,000
          |     75.40      24.60 |     100.0
```

```
   = 1 if |
  payment |
  late by |          probit
 90+ days |         0          1 |     Total
----------+----------------------+----------
        0 |       529         93 |       622
          |     52.90       9.30 |     62.20
----------+----------------------+----------
        1 |       216        162 |       378
          |     21.60      16.20 |     37.80
----------+----------------------+----------
    Total |       745        255 |     1,000
          |     74.50      25.50 |    100.00
```

(h) The predictive success for the last 500 observations after using the first 500 to estimate the model is given in the following tables. Using the 0.5 threshold, the out of sample predictive successes of the two models are similar to those for the whole sample.

= 1 if payment late by 90+ days	linear probability model 0	1	Total
0	257 51.40	37 7.40	294 58.80
1	130 26.00	76 15.20	206 41.20
Total	387 77.40	113 22.60	500 100.00

= 1 if payment late by 90+ days	probit 0	1	Total
0	254 50.80	40 8.00	294 58.80
1	128 25.60	78 15.60	206 41.20
Total	382 76.40	118 23.60	500 100.00

Let us investigate the success of using a 0.8 threshold and a 0.2 threshold. In both cases the threshold 0.8 is too high, with each model predicting only 3 defaults when there were 206 in the group of 500 loans.

= 1 if payment late by 90+ days	LPM with 0.8 threshold 0	1	Total
0	294 58.80	0 0.00	294 58.80
1	203 40.60	3 0.60	206 41.20
Total	497 99.40	3 0.60	500 100.00

= 1 if payment late by 90+ days	probit with 0.8 threshold 0	1	Total
0	294 58.80	0 0.00	294 58.80
1	203 40.60	3 0.60	206 41.20
Total	497 99.40	3 0.60	500 100.00

Now we try a threshold of 0.2. It is too low, predicting a very large number of defaults.

```
   = 1 if |
  payment |
  late by |          linear probability model with 0.2 threshold
 90+ days |          0          1 |     Total
-----------+----------------------+----------
         0 |         65        229 |       294
           |      13.00      45.80 |     58.80
-----------+----------------------+----------
         1 |         17        189 |       206
           |       3.40      37.80 |     41.20
-----------+----------------------+----------
     Total |         82        418 |       500
           |      16.40      83.60 |    100.00
```

```
   = 1 if |
  payment |
  late by |          probit with threshold 0.2
 90+ days |          0          1 |     Total
-----------+----------------------+----------
         0 |         87        207 |       294
           |      17.40      41.40 |     58.80
-----------+----------------------+----------
         1 |         21        185 |       206
           |       4.20      37.00 |     41.20
-----------+----------------------+----------
     Total |        108        392 |       500
           |      21.60      78.40 |    100.00
```

When making a loan there are two types of errors one can make. Give a loan to a person who defaults, meaning the model predicted a low probability of default but the person did default. In the table above, 21 such loans were given. The cost of such bad loans is that the lender must then take action (lawyers and the like) against the borrower to recover their funds. The other error is not giving a loan (presumably because the predicted probability of default was high) to someone who did not default. In the table above there were 207 of these errors. The cost of these mistakes is foregone profit. Lenders must find a threshold that minimizes the combined costs of these errors. It will be very institution specific, and we cannot evaluate what is optimal because we do not know the actual cost of each type of error.

EXERCISE 16.21

(a) The variable *NETPRICE* shows variation across the alternative brands, whereas *INCOME* is a household variable and is the same for all 4 alternatives on any choice occasion. Note that, in the following data for the first two households, the prices of the alternatives change within each group of 4 observations, but that income is constant.

```
    | hhid             alt   netprice    income |
    |-----------------------------------------|
 1. |    1     Skist-water        .79      47.5 |
 2. |    1       Skist-oil        .79      47.5 |
 3. |    1   ChiSea-water        .58      47.5 |
 4. |    1     ChiSea-oil        .58      47.5 |
    |-----------------------------------------|
 5. |    2     Skist-water        .56      47.5 |
 6. |    2       Skist-oil        .56      47.5 |
 7. |    2   ChiSea-water        .79      47.5 |
 8. |    2     ChiSea-oil        .79      47.5 |
```

(b) The choices among the 1500 cases follow. We observe that this group of consumers has a preference for tuna packed in water.

Alternatives summary for alt

Alternative value	label	Cases present	Frequency selected	Percent selected
1	Skist-water	1500	548	36.53
2	Skist-oil	1500	291	19.40
3	ChiSea-water	1500	475	31.67
4	ChiSea-oil	1500	186	12.40

(c) The probability that individual *i* chooses alternative *j*, for each of these 4 alternatives, is facilitated by using some simplifying notation. Let the variables and parameters for each alternative be denoted as follows:

$$xb(Skist\text{-}water)=\left(\beta_2 NETPRICE_{Skist\text{-}water}+\beta_3 DISPLAY_{Skist\text{-}water}+\beta_4 FEATURE_{Skist\text{-}water}\right)$$

$$xb(Skist\text{-}oil)=\left(\beta_{12}+\beta_2 NETPRICE_{Skist\text{-}oil}+\beta_3 DISPLAY_{Skist\text{-}oil}+\beta_4 FEATURE_{Skist\text{-}oil}\right)$$

$$xb(ChiSea\text{-}water)=\left(\beta_{13}+\beta_2 NETPRICE_{ChiSea\text{-}water}+\beta_3 DISPLAY_{ChiSea\text{-}water}+\beta_4 FEATURE_{ChiSea\text{-}water}\right)$$

$$xb(ChiSea\text{-}oil)=\left(\beta_{14}+\beta_2 NETPRICE_{ChiSea\text{-}oil}+\beta_3 DISPLAY_{ChiSea\text{-}oil}+\beta_4 FEATURE_{ChiSea\text{-}oil}\right)$$

Each variable should have a subscript "*i*" to denote the individual, but this has been suppressed to simplify notation. Each of the options has an intercept parameter except for Starkist-in-Water, which has none and serves as our base case. Then the probabilities that each of the options is chosen are:

$$p_{Skist\text{-}water}=\frac{\exp\left(xb(Skist\text{-}water)\right)}{\exp\left(xb(Skist\text{-}water)\right)+\exp\left(xb(Skist\text{-}oil)\right)+\exp\left(xb(ChiSea\text{-}water)\right)+\exp\left(xb(ChiSea\text{-}oil)\right)}$$

$$p_{Skist\text{-}oil}=\frac{\exp\left(xb(Skist\text{-}oil)\right)}{\exp\left(xb(Skist\text{-}water)\right)+\exp\left(xb(Skist\text{-}oil)\right)+\exp\left(xb(ChiSea\text{-}water)\right)+\exp\left(xb(ChiSea\text{-}oil)\right)}$$

$$p_{Chisea\text{-}water}=\frac{\exp\left(xb(Chisea\text{-}water)\right)}{\exp\left(xb(Skist\text{-}water)\right)+\exp\left(xb(Skist\text{-}oil)\right)+\exp\left(xb(ChiSea\text{-}water)\right)+\exp\left(xb(ChiSea\text{-}oil)\right)}$$

$$p_{Chisea\text{-}oil}=\frac{\exp\left(xb(Chisea\text{-}oil)\right)}{\exp\left(xb(Skist\text{-}water)\right)+\exp\left(xb(Skist\text{-}oil)\right)+\exp\left(xb(ChiSea\text{-}water)\right)+\exp\left(xb(ChiSea\text{-}oil)\right)}$$

(d) In the estimates that follow (obtained with Stata 11.1), we note that the estimated coefficients are all statistically significant, with the coefficient of the continuous variable *NETPRICE* carrying a negative sign and the indicator variables *DISPLAY* and *FEATURE* having positive signs. The alternative-specific variables are negative and statistically significant. From the probabilities in the previous part of the question, we see that all else being equal, the Starkist in oil, and Chicken of the Sea brands have a lower estimated probability of being selected than Starkist in water.

```
Alternative-specific conditional logit         Number of obs      =       6000
Case variable: hhid                            Number of cases    =       1500

Alternative variable: alt                      Alts per case: min =          4
                                                              avg =        4.0
                                                              max =          4

                                                  Wald chi2(3)    =     405.15
Log likelihood = -1537.2704                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
alt          |
    netprice |  -9.971961   .8628894   -11.56   0.000    -11.66319   -8.280729
     display |   1.635486   .2425727     6.74   0.000     1.160052    2.110919
     feature |   1.343511   .1366656     9.83   0.000     1.075652    1.611371
-------------+----------------------------------------------------------------
Skist_water  |  (base alternative)
-------------+----------------------------------------------------------------
Skist_oil    |
       _cons |  -.5959682   .0732714    -8.13   0.000    -.7395775   -.4523589
-------------+----------------------------------------------------------------
ChiSea_water |
       _cons |  -.5333423   .0816866    -6.53   0.000     -.693445   -.3732396
-------------+----------------------------------------------------------------
ChiSea_oil   |
       _cons |  -1.439991   .1002377   -14.37   0.000    -1.636453   -1.243529
```

(e) The marginal effects are given in the tables below. The first table gives the marginal effect of a price change for each of the brands on the probability of choosing Starkist in water. The "own" price effect is given using equation (16.24)

$$\frac{\partial p_{ij}}{\partial PRICE_{ij}} = p_{ij}\left(1 - p_{ij}\right)\beta_2$$

For example, given that *DISPLAY* and *FEATURE* are zero, the probabilities reduce to a dependence on the alternative specific constants and *NETPRICE*. If we set *NETPRICE* at its mean for each brand we can compute the probabilities of each choice being selected. For example, the first table shows that the probability of Starkist in water being selected is 0.406 with the price of each variable at its mean shown in the final column labeled "X".

The marginal effect of an increase in the net price of Starkist in water on the probability of choosing Startkist in water is

$$\frac{\partial p_{i1}}{\partial PRICE_{i1}} = p_{i1}\left(1 - p_{i1}\right)\beta_2 = .40557918 \times \left(1 - .40557918\right) \times -9.971961 = -2.40409$$

The change in probability is for a $1.00 change in price, which is more than the cost of the item. If the change is 10 cents, then we anticipate a reduction in the probability of purchase of 0.24.

The "cross-price" effect of a change in the price of one brand on the probability of selecting another brand is given by

$$\frac{\partial p_{ij}}{\partial PRICE_{ik}} = -p_{ij}\,p_{ik}\beta_2$$

The marginal effect of an increase in the price of Starkist in water on the probability of choosing Chicken of the Sea in water is, as shown in the third table below

$$\frac{\partial p_{i3}}{\partial PRICE_{i1}} = -p_{i3}\,p_{i1}\beta_2 = -.26646827 \times .40557918 \times -9.971961 = 1.07771$$

Pr(choice = Skist-water|1 selected) = .40557918

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
netprice							
Skist_water	-2.40409	.211831	-11.35	0.000	-2.81927	-1.98891	.68112
Skist_oil	.899315	.096762	9.29	0.000	.709665	1.08896	.68163
ChiSea_water	1.07771	.102948	10.47	0.000	.875935	1.27948	.66976
ChiSea_oil	.427063	.046209	9.24	0.000	.336495	.51763	.67167

Pr(choice = Skist-oil|1 selected) = .22235943

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
netprice							
Skist_water	.899315	.096762	9.29	0.000	.709664	1.08897	.68112
Skist_oil	-1.72431	.165592	-10.41	0.000	-2.04886	-1.39976	.68163
ChiSea_water	.590856	.060676	9.74	0.000	.471934	.709778	.66976
ChiSea_oil	.234138	.026839	8.72	0.000	.181533	.286742	.67167

Pr(choice = ChiSea-water|1 selected) = .26646827

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
netprice							
Skist_water	1.07771	.102948	10.47	0.000	.875935	1.27948	.68112
Skist_oil	.590856	.060675	9.74	0.000	.471934	.709778	.68163
ChiSea_water	-1.94915	.177526	-10.98	0.000	-2.29709	-1.6012	.66976
ChiSea_oil	.280583	.034964	8.02	0.000	.212055	.349111	.67167

Pr(choice = ChiSea-oil|1 selected) = .10559312

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
netprice							
Skist_water	.427063	.046209	9.24	0.000	.336495	.517631	.68112
Skist_oil	.234138	.02684	8.72	0.000	.181533	.286743	.68163
ChiSea_water	.280583	.034964	8.02	0.000	.212055	.349111	.66976
ChiSea_oil	-.941784	.099314	-9.48	0.000	-1.13644	-.747131	.67167

(f) Adding the individual specific variable *INCOME* to the model adds 3 parameters to estimate. Like alternative specific constants, coefficients of individual specific variables are different for each alternative, with Starkist in water again set as the base case. The estimates are

```
Alternative-specific conditional logit         Number of obs      =       6000
Case variable: hhid                            Number of cases    =       1500

Alternative variable: alt                      Alts per case: min =          4
                                                              avg =        4.0
                                                              max =          4

                                                  Wald chi2(6)    =     419.76
Log likelihood = -1529.3439                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
alt          |
    netprice |   -9.99618   .8641534   -11.57   0.000    -11.68989   -8.302471
     display |   1.619318   .2429992     6.66   0.000     1.143048    2.095587
     feature |   1.336417   .1367137     9.78   0.000     1.068463    1.604371
-------------+----------------------------------------------------------------
Skist_water  |  (base alternative)
-------------+----------------------------------------------------------------
Skist_oil    |
      income |   -.021638   .0060061    -3.60   0.000    -.0334097   -.0098662
       _cons |  -.0673146   .1611304    -0.42   0.676    -.3831243    .2484952
-------------+----------------------------------------------------------------
ChiSea_water |
      income |  -.0027101   .0058179    -0.47   0.641     -.014113    .0086928
       _cons |  -.4607403   .1710712    -2.69   0.007    -.7960337   -.1254469
-------------+----------------------------------------------------------------
ChiSea_oil   |
      income |   -.012768   .0075311    -1.70   0.090    -.0275287    .0019926
       _cons |  -1.117534   .2119356    -5.27   0.000     -1.53292   -.7021478
```

The likelihood ratio test is based on the difference in the log-likelihood values for the two models.

$$LR = 2\left(\ln L_U - \ln L_R\right) = 2\left(-1529.3439 - \left(-1537.2704\right)\right) = 15.853$$

The test critical value is the 95th percentile of the $\chi^2_{(3)}$ distribution, which is 7.815. Thus we reject the hypothesis that the coefficients on *INCOME* are all zero, and conclude that *INCOME* has an effect on these choices.

(g) The marginal effects of *NETPRICE*, using the specified values for *DISPLAY*, *FEATURE* and *INCOME*, and with *NETPRICE* at its mean for each brand, are as follows:

```
Pr(choice = Skist-water|1 selected) = .41970781
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
netprice     |
 Skist_water | -2.4346    .213621  -11.40   0.000  -2.85329  -2.01591   .68112
   Skist_oil | .855809    .095308    8.98   0.000   .669008   1.04261   .68163
ChiSea_water |  1.1472    .110758   10.36   0.000   .930117   1.36428   .66976
  ChiSea_oil | .431595    .048578    8.88   0.000   .336383   .526806   .67167
```

```
Pr(choice = Skist-oil|1 selected) = .20398375
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
netprice     |
 Skist_water | .855809    .095308    8.98   0.000   .669008   1.04261   .68112
   Skist_oil |-1.62312    .161558  -10.05   0.000  -1.93977  -1.30648   .68163
ChiSea_water | .557554    .059582    9.36   0.000   .440775   .674332   .66976
  ChiSea_oil | .209761    .025517    8.22   0.000   .159749   .259773   .67167
```

```
Pr(choice = ChiSea-water|1 selected) = .27343699
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
netprice     |
 Skist_water |  1.1472    .110758   10.36   0.000   .930116   1.36428   .68112
   Skist_oil | .557554    .059582    9.36   0.000   .440776   .674332   .68163
ChiSea_water |-1.98593    .181947  -10.91   0.000  -2.34254  -1.62932   .66976
  ChiSea_oil | .281181    .036704    7.66   0.000   .209243   .353119   .67167
```

```
Pr(choice = ChiSea-oil|1 selected) = .10287145
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
netprice     |
 Skist_water | .431595    .048578    8.88   0.000   .336383   .526806   .68112
   Skist_oil | .209761    .025517    8.22   0.000   .159749   .259773   .68163
ChiSea_water | .281181    .036704    7.66   0.000   .209243   .353119   .66976
  ChiSea_oil |-.922537     .10131   -9.11   0.000   -1.1211  -.723972   .67167
```

EXERCISE 16.23

(a) Of the 4,483 individuals, 1777 (39.64%) choose $HSAT3 = 1$; 1792 (39.97%) choose $HSAT3 = 2$; and 914 (20.39%) choose $HSAT3 = 3$.

(b) The estimates are given in the following table. All the coefficients are significant at the 5% level or better.

Table XR16.23 Ordinal probit estimates

AGE	-0.05571^{***}	(-4.241)
AGE^2	0.00041^{**}	(2.722)
EDUC2	0.04785^{***}	(6.464)
WORKING	0.09107^{*}	(2.364)
$\tilde{\mu}_1$	-1.25283^{***}	(-4.410)
$\tilde{\mu}_2$	-0.11918	(-0.420)
N	4483	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(c) Let **xb** denote the coefficients and variable values

$$\mathbf{xb} = \tilde{\beta}_1 AGE + \tilde{\beta}_2 AGE^2 + \tilde{\beta}_3 EDUC2 + \tilde{\beta}_4 WORKING$$

In this problem $\mathbf{xb} = \tilde{\beta}_1 40 + \tilde{\beta}_2 40^2 + \tilde{\beta}_3 16 + \tilde{\beta}_4 = -0.71968109$

$$\frac{\partial \hat{P}[HSAT3 = 1]}{\partial AGE} = -\phi(\tilde{\mu}_1 - \mathbf{xb}) \times \left[\tilde{\beta}_1 + 2\tilde{\beta}_2 AGE\right] = 0.00799711$$

$$\frac{\partial \hat{P}[HSAT3 = 2]}{\partial AGE} = \left[\phi(\tilde{\mu}_1 - \mathbf{xb}) - \phi(\tilde{\mu}_2 - \mathbf{xb})\right] \times \left[\tilde{\beta}_1 + 2\tilde{\beta}_2 AGE\right] = -0.0002996$$

$$\frac{\partial \hat{P}[HSAT3 = 3]}{\partial AGE} = \phi(\tilde{\mu}_2 - \mathbf{xb}) \times \left[\tilde{\beta}_1 + 2\tilde{\beta}_2 AGE\right] = -0.0076975$$

We estimate that an increase in age increases the probability of *HSAT3* = 1, it reduces the probability of *HSAT3* = 2, and it reduces the probability of *HSAT3* = 3 even more.

(d) In this problem $\mathbf{xb} = \tilde{\beta}_1 70 + \tilde{\beta}_2 70^2 + \tilde{\beta}_3 16 + \tilde{\beta}_4 = -1.0461234$.

$$\frac{\partial \hat{P}[HSAT3 = 1]}{\partial AGE} = -\phi(\tilde{\mu}_1 - \mathbf{xb}) \times \left[\tilde{\beta}_1 + 2\tilde{\beta}_2 AGE\right] = -0.00052496$$

$$\frac{\partial \hat{P}[HSAT3 = 2]}{\partial AGE} = \left[\phi(\tilde{\mu}_1 - \mathbf{xb}) - \phi(\tilde{\mu}_2 - \mathbf{xb})\right] \times \left[\tilde{\beta}_1 + 2\tilde{\beta}_2 AGE\right] = 0.00017596$$

$$\frac{\partial \hat{P}[HSAT3 = 3]}{\partial AGE} = \phi(\tilde{\mu}_2 - \mathbf{xb}) \times \left[\tilde{\beta}_1 + 2\tilde{\beta}_2 AGE\right] = 0.000349$$

(e) Working is a 0/1 variable and therefore differentiation is not feasible. Instead we can compute the differences in the probabilities directly. The probabilities of the various outcomes are

$$\hat{P}[HSAT3 = 1] = \Phi(\tilde{\mu}_1 - \mathbf{xb})$$
$$\hat{P}[HSAT3 = 2] = \Phi(\tilde{\mu}_2 - \mathbf{xb}) - \Phi(\tilde{\mu}_1 - \mathbf{xb})$$
$$\hat{P}[HSAT3 = 3] = 1 - \Phi(\tilde{\mu}_2 - \mathbf{xb})$$

Let $\mathbf{xb0} = \tilde{\beta}_1 AGE + \tilde{\beta}_2 AGE^2 + \tilde{\beta}_3 EDUC2 = -0.81075218$

And $\mathbf{xb1} = \tilde{\beta}_1 AGE + \tilde{\beta}_2 AGE^2 + \tilde{\beta}_3 EDUC2 + \tilde{\beta}_4 WORKING = -0.71968109$

$$\hat{P}\left[HSAT3 = 1 \,|\, \mathbf{xb1}\right] - \hat{P}\left[HSAT3 = 1 \,|\, \mathbf{xb0}\right] = \Phi\left(\tilde{\mu}_1 - \mathbf{xb1}\right) - \Phi\left(\tilde{\mu}_1 - \mathbf{xb0}\right) = -0.03225125$$

$$\begin{aligned}
&\hat{P}\left[HSAT3 = 2 \,|\, \mathbf{xb1}\right] - \hat{P}\left[HSAT3 = 2 \,|\, \mathbf{xb0}\right] \\
&\quad = \left[\Phi\left(\tilde{\mu}_2 - \mathbf{xb1}\right) - \Phi\left(\tilde{\mu}_1 - \mathbf{xb1}\right)\right] - \left[\Phi\left(\tilde{\mu}_2 - \mathbf{xb0}\right) - \Phi\left(\tilde{\mu}_1 - \mathbf{xb0}\right)\right] \\
&\quad = 0.00276819
\end{aligned}$$

$$\begin{aligned}
&\hat{P}\left[HSAT3 = 3 \,|\, \mathbf{xb1}\right] - \hat{P}\left[HSAT3 = 3 \,|\, \mathbf{xb0}\right] \\
&\quad = \left[1 - \Phi\left(\tilde{\mu}_2 - \mathbf{xb1}\right)\right] - \left[1 - \Phi\left(\tilde{\mu}_2 - \mathbf{xb0}\right)\right] = 0.02948306
\end{aligned}$$

EXERCISE 16.25

(a) The estimates are in column (1) of Table XR16.25a. As discussed on page 719, strategy 2, the OLS estimator using all the observations is biased and inconsistent.

Table XR16.25a

	(1) ols	(2) ols if > 0	(3) probit
C	-20.6750	1143.9304***	1.3900**
	(-0.108)	(4.665)	(3.126)
EXPER	42.8949***	28.8830***	0.0644***
	(11.847)	(6.457)	(9.101)
EDUC	31.1346*	-21.0209	
	(2.434)	(-1.332)	
HHOURS	-0.0341	0.0217	
	(-0.694)	(0.350)	
KIDSL6			-0.3938***
			(-4.083)
KIDS618			0.1478***
			(3.716)
MTR			-2.7550***
			(-4.442)
LARGECITY			-0.1618
			(-1.526)
N	753	428	753

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) These estimates are in Table XR16.25a, column (2). As discussed on page 719, strategy 1, the OLS estimator using all the observations is biased and inconsistent.

(c) The probit estimates are in Table XR16.25a, column (3). Experience has a positive and significant effect on the decision to join the labor force. Having children less than 6 years old has a negative and significant effect on the probability of joining the labor force, while having children in 6-18 range has a positive and significant effect on the probability of joining the labor force. The marginal tax rate facing the wife has a negative and significant effect on the probability of joining the labor force.

(d) The sample mean of $\tilde{\lambda}$ is 0.7163018 and the sample variance is 0.11729.

(e) The OLS estimates are in column (1) and conventional standard errors are in column (2) of Table XR16.25b. Note that none of the coefficients other than the constant term and IMR are significant in these models. These standard errors are incorrect because they ignore the presence of the variable IMR which was generated from a first stage equation.

Table XR16.25b

	(1) ols imr	(2) conventional	(3) HC3	(4) Boot
C	2094.6455***	(340.275)	(356.011)	(357.957)
EXPER	5.4648	(7.371)	(7.438)	(7.530)
EDUC	-25.8066	(15.560)	(15.333)	(16.016)
HHOURS	-0.0300	(0.062)	(0.066)	(0.065)
IMR	-797.6963***	(201.469)	(202.476)	(201.999)
N	428			

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(f) The robust standard errors using HC3 are in column (3) of Table XR16.25b. They are not much different from the conventional standard errors in column (2). These standard errors are incorrect because they ignore the presence of the variable IMR which was generated from a first stage equation.

(g) The bootstrap standard errors are in column (4) of Table XR16.25b. They are not much different from the other standard errors in columns (2) and (3). These standard errors are incorrect because they ignore the presence of the variable IMR which was generated from a first stage equation.

(h) Rather than report the estimates in a summary table we will report the full Stata 15.1 output for these models. The two-step estimates are on a separate page and the maximum likelihood estimates on another page.

The two step estimates are reported first, then the probit estimates used to construct the IMR. Note that Stata reports the constant term, _cons, last. Followed by the Mills lambda. Then followed by rho, which is an estimate of the correlation between the probit equation error and the equation of interest error. If rho is zero then there is no selection bias. The sigma is the estimated square root of the error variance for the equation of interest error.

Important notes:

1. It is key to note that based on these two step estimates we cannot reject the null hypothesis that the parameters of the equation of interest are zero. That is indicated by the Wald statistic 3.79 with *p*-value 0.2849.
2. Also note that the coefficient of the IMR is statistically significant, which is an indication that “selection bias” exists and therefore one of these Heckit, Heckman selection models is appropriate.
3. Note that the coefficient estimates are the same as in column (1) of Table XR16.25b, although the standard errors are different. These two step estimator standard errors are corrected for the fact that IMR is an estimated quantity.

In the maximum likelihood estimation procedure all the parameters are estimated jointly, and the MLE is the “best”, in a certain sense, estimator. It is more efficient than the two step estimator. In the estimated coefficients of the equation of interest we find that experience has a statistically significant coefficient and the Wald statistic now has a $p = 0.0044$. Estimates of rho, sigma and lambda are given at the bottom. Also reported at the bottom is the likelihood ratio test statistic for the null hypothesis that rho = 0. We reject that at the 1% level and conclude that correcting for selection bias is important in this model. You should ignore, for now, the /athrho and /lnsigma entries.

Two step estimates

```
Heckman selection model -- two-step estimates   Number of obs      =       753
(regression model with sample selection)              Selected     =       428
                                                      Nonselected  =       325

                                                Wald chi2(3)       =      3.79
                                                Prob > chi2        =    0.2849

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hours        |
       exper |   5.464807    8.34185     0.66   0.512    -10.88492    21.81453
        educ |  -25.80655   15.89106    -1.62   0.104    -56.95246    5.339357
      hhours |  -.0299916   .0624497    -0.48   0.631    -.1523907    .0924074
        cons |   2094.645   354.2201     5.91   0.000     1400.387    2788.904
-------------+----------------------------------------------------------------
lfp          |
       exper |   .0643837   .0070741     9.10   0.000     .0505187    .0782486
      kids16 |  -.3938219   .0964479    -4.08   0.000    -.5828564   -.2047875
     kids618 |   .1477796   .0397675     3.72   0.000     .0698368    .2257225
         mtr |  -2.754952   .6202055    -4.44   0.000    -3.970533   -1.539372
   largecity |  -.1617942   .1060387    -1.53   0.127    -.3696261    .0460378
       _cons |   1.389992   .4446666     3.13   0.002     .5184613    2.261522
-------------+----------------------------------------------------------------
/mills       |
      lambda |  -797.6963   219.3468    -3.64   0.000    -1227.608   -367.7845
-------------+----------------------------------------------------------------
         rho |   -0.86173
       sigma |  925.69693
```

Maximum likelihood estimates

```
Heckman selection model                         Number of obs     =        753
(regression model with sample selection)              Selected    =        428
                                                      Nonselected =        325

                                                Wald chi2(3)      =      13.09
Log likelihood = -3872.079                      Prob > chi2       =     0.0044

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hours        |
       exper |   17.29909    5.50024     3.15   0.002     6.518817    28.07936
        educ |  -24.05109    15.7628    -1.53   0.127    -54.94562    6.843433
      hhours |  -.0079731    .062278    -0.13   0.898    -.1300356    .1140895
       _cons |   1632.746    275.742     5.92   0.000     1092.302    2173.191
-------------+----------------------------------------------------------------
lfp          |
       exper |   .0645032   .0071135     9.07   0.000      .050561    .0784455
      kidsl6 |  -.4193251   .0934692    -4.49   0.000    -.6025213   -.2361288
     kids618 |   .1331413   .0395943     3.36   0.001     .0555379    .2107448
         mtr |  -3.242047   .6118144    -5.30   0.000    -4.441181   -2.042912
  largecity  |  -.1974712   .1018975    -1.94   0.053    -.3971865    .0022442
        cons |   1.765425   .4383326     4.03   0.000      .906309    2.624541
-------------+----------------------------------------------------------------
     /athrho |  -.5569817   .1504488    -3.70   0.000    -.8518558   -.2621075
    /lnsigma |   6.670743   .0493731   135.11   0.000     6.573974    6.767513
-------------+----------------------------------------------------------------
         rho |  -.5057345   .1119689                     -.6920378   -.2562657
       sigma |   788.9818   38.95449                      716.2102    869.1474
      lambda |  -399.0153   103.4982                      -601.868   -196.1627
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =     9.73   Prob > chi2 = 0.0018
```

EXERCISE 16.27

(a) The estimates are in Table XR16.27.

The coefficient of *FEMALE* is positive and significant, indicating the females have a higher expected number of doctor visits. Using the percentage change in the conditional mean $100\left[e^{\delta}-1\right]\%$, females have a 22.95% higher conditional mean number of visits.

SELF has a negative and significant coefficient suggesting that self-employed people make a lower expected number of doctor visits than those who are not self-employed. We estimate that they have a 33.06% lower conditional mean.

POST has a negative and significant coefficient suggesting that those with a post-secondary degree have fewer expected number of doctor visits than those without such a degree. We estimate that they have 17.77% lower conditional mean.

The coefficient of *PUBLIC* is positive and significant at the 5% level, indicating that the expected number of doctor visits for those who are insured by the public. We estimate those with public insurance will have 8.45% higher expected number of doctor visits.

Table XR16.27 Poisson Regression

	Coef.	*t*-value
C	2.9579***	(11.783)
FEMALE	0.2066***	(8.709)
AGE	-0.0456***	(-5.046)
AGE2	0.0006***	(6.084)
SELF	-0.4013***	(-6.888)
LINC	-0.1792***	(-7.186)
POST	-0.1956***	(-4.666)
PUBLIC	0.0811*	(1.997)
N	3000	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) The percentage change with respect to *AGE* is

$$100\frac{\partial E(DOCVIS\,|\,\mathbf{X})/E(DOCVIS\,|\,\mathbf{X})}{\partial AGE}=\frac{\%\Delta E(DOCVIS\,|\,\mathbf{X})}{\partial AGE}=100(\beta_2+2\beta_3 AGE)\%$$

For a 30 year old the estimated percentage change is −0.84%; for a 50 year old it is 1.63%; and for a 70 year old it is 4.10%.

(c) Write the model as follows, where **xb** represents everything else in the model,

$$\begin{aligned}E(DOCVIS\,|\,\mathbf{X})&=\exp(\beta\ln(INCOME)+\mathbf{xb})\\&=\exp(\beta\ln(INCOME))\exp(\mathbf{xb})\\&=INCOME^{\beta}\exp(\mathbf{xb})\end{aligned}$$

Then,

$$\frac{\partial E(DOCVIS\,|\,\mathbf{X})}{\partial INCOME}=\beta INCOME^{\beta-1}\exp(\mathbf{xb})$$

$$\Rightarrow$$

$$\frac{100}{100}\frac{\partial E(DOCVIS\,|\,\mathbf{X})}{\partial INCOME}\frac{INCOME}{E(DOCVIS\,|\,\mathbf{X})}=\frac{\%\Delta E(DOCVIS\,|\,\mathbf{X})}{\%\Delta INCOME}=\beta$$

Thus the coefficient of *LINC* is the elasticity of expected doctor visits with respect to *INCOME*. We estimate that a 1% increase in income reduces the expected number of doctor visits by 0.179%.

(d) The percentage of correct predictions is 12.57% within the sample of 3000 observations. For the 979 other observations, not used in the estimation, the percentage of correct predictions is 14.3%.

(e) The percentage of correct predictions is 45.5% within the sample of 3000 observations. For the 979 other observations, not used in the estimation, the percentage of correct predictions is 45.25%.

EXERCISE 16.29

(a) These estimates are in column (1) of Table XR16.29a. The variable *ADDON*'s coefficient is not statistically significant.

Table XR16.29a

	(1) part (a)	(2) part (b)	(3) part (c)	(4) part (d)
C	1.2842^{***}	-0.0025	1.1906^{**}	1.1906^{**}
	(3.400)	(-0.192)	(3.141)	(2.690)
AGE	0.0557^{***}	0.0003	0.0603^{***}	0.0603^{***}
	(7.684)	(1.229)	(8.060)	(7.163)
FEMALE	0.3586^{*}	0.0016	0.4358^{*}	0.4358^{*}
	(2.023)	(0.278)	(2.381)	(2.065)
WORKING	-0.4642^{*}	-0.0177^{**}	-0.5690^{**}	-0.5690^{*}
	(-2.383)	(-2.724)	(-2.936)	(-2.533)
HHNINC2	-0.0002^{***}	0.0000^{***}	-0.0001	-0.0001
	(-5.105)	(3.545)	(-1.115)	(-0.928)
ADDON	0.3602			-17.0774
	(0.891)			(-1.910)
WHITEC		0.0220^{***}		
		(3.494)		
SELF		0.0251^{*}		
		(2.005)		
ADDON_P			-17.0774^{*}	
			(-2.243)	
N	4483	4483	4483	4483
ln(*L*)	-13640.9357	1610.7869	-13638.9834	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) These estimates are in column (2) of Table XR16.29a. We see from the tabled results that the coefficients of *WHITEC* and *SELF* are significant. The *F*-test statistic value for their joint significance is 7.39 with $p = 0.0006$, that is to say very significant. However, the value is less than "10" the rule of thumb value for the success of IV estimation with an endogenous variable.

(c) These estimates are in column (3) of Table XR16.29a. In the table the fitted value of *ADDON* is *ADDON_P*. The coefficient of the fitted variable is negative and significant, suggesting that those insured by add-on insurance are estimated to have a lower expected number of doctor visits.

(d) These estimates are in column (4) of Table XR16.29a. The coefficient estimates are the same as in part (c), as expected, but the standard errors are different. Using the correct standard errors means the estimated coefficient of *ADDON* is no longer significant at the 5% level.

(e) The first stage estimates are in column (1) of Table XR16.29b. The two proposed IV, *WHITEC* and *SELF* are both significant. The second stage regression is in column (2) of Table XR16.29b. The coefficient of *PHAT* is negative and significant. The results are not the same as in part (d) because *PHAT* from the probit model will be different from *ADDON_P* from the linear probability model.

Table XR16.29b

	(1) probit	(2) 2sls	(3) IV
C	-2.3313^{***} (-11.565)	1.2226^{**} (3.215)	1.1799^{*} (2.569)
AGE	0.0040 (1.118)	0.0587^{***} (7.975)	0.0608^{***} (6.937)
FEMALE	0.0330 (0.396)	0.4231^{*} (2.350)	0.4446^{*} (2.053)
WORKING	-0.2975^{**} (-2.702)	-0.5583^{**} (-2.848)	-0.5810^{*} (-2.478)
HHNINC2	0.0001^{***} (3.856)	-0.0001 (-1.428)	-0.0001 (-0.651)
ADDON			-19.0712 (-1.886)
PHAT		-13.7663^{*} (-2.372)	
WHITEC	0.3650^{***} (3.578)		
SELF	0.3866^{**} (2.626)		
N	4483	4483	4483
ln(*L*)	-580.6687	-13639.2094	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(f) The IV estimates are in column (3) of Table XR16.29b. The coefficient of *ADDON* is negative but not significant at 5%. The first stage *F* is 7.77, so again we may face a weak instrument problem. The estimates are different from those in part (e) because the first stage regression is different.