14.32 Pset 4 Solutions

due November 7, 2023

1. Question 1

   a) We first calculate the probability of graduating for the student who studies 10 hours:

$$\begin{aligned} P(grad = 1 | hsGPA = 3.0, SAT = 1200, study = 10) \\ &= \Phi(-0.73 + 0.15 \cdot 3 + 0.00036 \cdot 1200 + 0.046 \cdot 10) \\ &= \Phi(0.612) = 0.7297 \end{aligned}$$

   The probability of graduating for the student who studies 5 hours is

$$\begin{aligned} P(grad = 1 | hsGPA = 3.0, SAT = 1200, study = 5) \\ &= \Phi(-0.73 + 0.15 \cdot 3 + 0.00036 \cdot 1200 + 0.046 \cdot 5) \\ &= \Phi(0.382) = 0.6487 \end{aligned}$$

   Thus, the estimated difference in graduation probability is $0.7297 - 0.6487 = 0.081$, or 8.1%.

   b)

$$\frac{\partial P(employed = 1 | age, educ)}{\partial age} = \frac{\partial \Phi(age, educ)}{\partial age} = f(age, educ)(\beta + 2\gamma age)$$

   Using the formula for standard normal pdf, we have:

$$= \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-(\alpha + \beta age + \gamma age^2 + \delta educ)^2}{2}\Big)\Big(\beta + 2\gamma age\Big)$$

   c) No. The response variable only takes the values of 0 and 1, while the predicted values are probabilities anywhere between 0 and 1 (and sometimes even greater or smaller). Thus, the size of the residuals of a linear probability model will grow or shrink as the predicted values grow or shrink, and their distribution is incomparable to a normal or logistic distribution.

2. Question 2

   a) The p-value of our F-test is 0.1915. Thus, the variables are not jointly significant at the 5% level.

b) The p-value of our likelihood-ratio test is 0.1785. Thus, we cannot say that any of the added variables help to predict participation in the training program.

c) Based on our results, we can say that participation in job training *can* be treated as exogenous. It does not appear to be explained in any part by other variables in the dataset.

d) Based on our linear probability model, participating in training decreases the probability of being unemployed in 1978 by 11.06 percentage points. This finding is statistically significant, given by the t value of -2.5.

e) The resulting equation is

$$P(unemp78_i|train_i) = \Phi(-.375 - 0.321 \cdot train_i)$$

We cannot compare the coefficient of the probit model with the coefficient of the linear model. This is because the linear coefficient can be interpreted as the marginal effect of *train* on the probability of being employed, while the probit coefficient cannot.

f) The fitted probabilities are equal because we have only one independent variable that takes on the value of 0 or 1. Thus, we can directly solve for the probability of unemployment when $train = 0$ and when $train = 1$ for both models.

Linear model:

$$P(unemp78_i = 1|train_i = 0) = \mathbf{0.3538}$$
$$P(unemp78_i = 1|train_i = 1) = 0.3538 - 0.1106 = \mathbf{0.2432}$$

Probit model:

$$P(unemp78_i = 1|train_i = 0) = \Phi(-0.374) = \mathbf{0.3538}$$
$$P(unemp78_i = 1|train_i = 1) = \Phi(-0.374 - 0.321) = \Phi(-0.696) = \mathbf{0.2432}$$

g) After adding controls, the fitted probabilities are no longer identical.

h) The marginal effects of the linear and probit models are quite similar, both weighing *train* and *black* the most. However, the probit model weights *train* slightly more than the linear model.

i) The marginal effects of the logit model are very similar to the linear and probit. It is a bit closer to the linear model, but generally, the partial effects given by all three models are incredibly close.

Table of Marginal Effects by Variable

Variable	Linear model	Probit model	Logit model
train	-.11170278	-.11317326	-.11233012
	.04430607	.04332547	.04353486
unem74	.03869256	.03567252	.03977036
	.07159545	.07142978	.07145149
unem75	.01596126	.02138872	.01679112
	.06673007	.06623926	.0658879
age	.00004332	.00022721	.0000735
	.00315477	.00306672	.0030787
educ	.00014424	-.00063601	-.00032145
	.01236873	.01237116	.01228984
black	.18883279	.21306164	.22386986
	.08133816	.09088463	.10019642
hisp	-.03770107	-.05545892	-.04947632
	.10870003	.12736293	.14083644
married	-.0254373	-.02614833	-.02758009
	.0596735	.05953388	.06013886

Legend: b/se