Question 1. For the population of men who grew up with disadvantaged backgrounds, let poverty be a dummy variable equal to one if a man in currently living below the poverty line, and zero otherwise. The variable *age* is an individual's age and *educ* is total years of schooling. Let *vocat* be an indicator equal to unity if a man's high school offered vocational training. Using a random sample of 850 men, you obtain

$$Pr(poverty = 1 \mid educ, age, vocat) = \Lambda(.453 - .016age - .087educ - .049vocat)$$

where $\Lambda(z) = \exp(z)/(1 + \exp(z))$ is the logistic function.

(a) For a 40-year old man with 12 years of education, what is the estimated effect of having vocational training available in high school on the probability of currently living in poverty?

$$Z = 0.453 - 0.016age - 0.87educ - 0.49vocat$$

$$\begin{aligned} Z_{(age=40,educ=12,vocat=1)} &= 0.453 - 0.016 \cdot 40 - 0.087 \cdot 12 - 0.049 \cdot 1 \\ &= 0.453 - 0.64 - 1.044 - 0.049 \\ &= -1.28 \\ Z_{(age=40,educ=12,vocat=0)} &= 0.453 - 0.016 \cdot 40 - 0.087 \cdot 12 - 0.049 \cdot 0 \\ &= 0.453 - 0.64 - 1.044 \\ &= -1.231 \end{aligned}$$

$$\begin{aligned} \frac{\Delta Pr(poverty = 1 \mid educ, age, vocat)}{\Delta vocat} &= \Lambda\left(Z_{(age=40,edtc=12,vocat=1)}\right) - \Lambda\left(Z_{(age=40,edtc=12,vocat=0)}\right) \\ &= \Lambda(-1.28) - \Lambda(-1.231) \\ &= \frac{\exp(-1.28)}{1 + \exp(-1.28)} - \frac{\exp(-1.231)}{1 + \exp(-1.231)} \\ &= -0.00845 \end{aligned}$$

(b) A new welfare program for "out-of-poverty by education" is designed to foster longer support in education from elementary school (K-6) to high school (K-12). What will be the partial effect of the average individual affected by the welfare package for education?

$$\begin{aligned} \widehat{PEA} &= \Lambda(0.453 - 0.016\,\overline{age} - 0.087\,\overline{educ} - 0.049\,\overline{vocat}) \cdot (-0.087) \\ &= \frac{\exp(0.453 - 0.016\,\overline{age} - 0.087\,\overline{educ} - 0.049\,\overline{vocat})}{1 + \exp(0.453 - 0.016\,\overline{age} - 0.087\,\overline{educ} - 0.049\,\overline{vocat})} \cdot (-0.087) \end{aligned}$$

(c) What about average partial effect of all individuals?

$$\begin{aligned} \widehat{APE} &= \frac{1}{n}\sum_{i=1}^{n} g\left(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k \chi_k\right) \cdot \hat{\beta}_j \\ &= \frac{1}{850}\sum_{i=1}^{850} g(0.453 - 0.016age - 0.87educ - 0.49rocat) \cdot (-0.87) \end{aligned}$$

(d) Given (c), what's the form of average treatment effect?

Given (c), ATE is same as APE.

(e) Are the effect constant over varying age, educ, vocat? What will be the sign of coefficient, if you add age^2 in the logistic regression? How can you justify the sign? What about $educ^2$?

The effect is not constant over varying age, educ, vocat.
If we add age^2 in the logistic regression, the sign of age^2's coefficient will be positive. As he gets older, the probability of poverty decreases at first, and then at some point, it will grow.
If we add $educ^2$ in the logistic regression, the sign of $educ^2$'s coefficient will be negative. Because if he had a proper job education, the probability of poverty would decrease.

Question 2. The following estimates were calculated using a sample of 7,634 women respondents from the General Household Survey 1995. The dependent variable takes the value 1 if the woman was in paid employment, and 0 otherwise. where *high* is 1 if the respondent has a higher educational qualification, 0 otherwise;

	OLS	Logit	Probit
high	0.093	0.423	0.259
	(0.015)	(0.071)	(0.043)
noqual	−0.210	−0.898	−0.554
	(0.013)	(0.056)	(0.035)
age	0.038	0.173	0.108
	(0.003)	(0.124)	(0.008)
age2	−0.051	−0.230	−0.142
	(0.003)	(0.069)	(0.009)
mar	0.024	0.103	0.063
	(0.009)	(0.057)	(0.035)
Constant	−0.068	−2.587	−1.593
	(0.049)	(0.225)	(0.137)

noqual is 1 if the respondent has no qualification, 0 otherwise; *age* is the age in years; *age*2 is (age × age) /100; *mar* is 1 is married, 0 otherwise. Conventionally calculated standard errors are in brackets for the ordinary least squares (OLS) results and asymptotic standard errors are in brackets elsewhere.

(a) Explain how the Probit estimates are obtained when the model has no intercept and there is only one explanatory variable.
The slope becomes sharper when the Probit has no intercept since the sign of constant is negative.

(b) Using all three sets of estimates, test the null hypothesis that the coefficient of *mar* is zero. Which test statistic(s) would you consider more reliable? Explain

The hypothesis

$$\begin{cases} H_0 : \beta_{mar} = 0 \\ H_1 : \beta_{mar} \neq 0 \end{cases}$$

$$\text{OLS: } Z = \frac{\hat{\beta}_{mar}}{SE\left(\hat{\beta}_{mar}\right)} = \frac{0.024}{0.009} = 2.67 > 1.96 \Rightarrow \text{ Reject } H_0$$

$$\text{Logit: } Z = \frac{\hat{\beta}_{mar}}{SE\left(\hat{\beta}_{mar}\right)} = \frac{0.103}{0.057} = 1.80 < 1.96 \Rightarrow \text{ Accept } H_0$$

$$\text{Probit: } Z = \frac{\hat{\beta}_{mar}}{SE\left(\hat{\beta}_{mar}\right)} = \frac{0.063}{0.035} = 1.8 < 1.96 \Rightarrow \text{ Accept Ho}$$

If data is normal distributed, Probit test statistic is more reliable.

(c) Using the OLS and Probit estimates calculate the estimated probability of being in paid employment for a married woman, aged 40 with a higher education qualification. Comment on your results.

$$\begin{aligned}
\text{OLS} : & Pr(high = 1, age = 40, age2 = 16, mar = 1) \\
& = 0.093 \cdot 1 + 0.038 \cdot 40 - 0.051 \cdot 16 - 0.024 \cdot 1 - 0.068 \\
& = 0.093 + 1.52 - 0.816 - 0.024 - 0.068 \\
& = 0.753 \\
\text{Probit} : & Pr(high = 1, age = 40, age2 = 16, mar = 1) \\
& = \Phi(0.259 \cdot 1 + 0.108 \cdot 40 - 0.142 \cdot 16 + 0.063 \cdot 1 - 1.593) \\
& = \Phi(0.259 + 4.32 - 2.272 + 0.063 - 1.593) \\
& = \Phi(0.777) \\
& = 0.781
\end{aligned}$$

Probit is more reliable. Because LPM have three shortcomings. (It is described in the following question.)

(d) Why is OLS (or Linear Probability Model) an inferior option to the other two models? Statisticians often claim that Logit and Probit are practically equivalent in real world applications. Given above, can you support the claim?

The reason why OLS(LPM) is an inferior option to the other two models:
① The fitted values from an OLS regression are never guaranteed to lie between zero and one.
② The estimated partial effects are constant; may lead to silly estimated effects for large changes.[Quadratic terms are typically limited]
③ The LPM exhibits heteroskedasticity - A4 violation (not efficient)

Except for the extreme point, Probit exaggerates the probability a little more than Logit. And Logit converge faster and easier to calculate than Probit, so they cost less to compute. If Logit and Probit are practically equivalent in real world, I think it is better to use Logit, which costs less to compute.