1. (a) Estimated effect of vocat $= \Lambda(0.453 - 0.016 \times 40 - 0.087 \times 12 - 0.049 \times 1)$

$$- \Lambda(0.453 - 0.016 \times 40 - 0.087 \times 12)$$

$$= \Lambda(X_1) - \Lambda(X_2) = \frac{e^{X_1}}{1+e^{X_1}} - \frac{e^{X_2}}{1+e^{X_2}}$$

$X_1 = -1.28$, $X_2 = -1.231$

$$= \frac{e^{-1.28}}{1+e^{-1.28}} - \frac{e^{-1.231}}{1+e^{-1.231}} = -8.456 \times 10^{-3}$$

$$\fallingdotseq -0.0085$$

(b) $$\widehat{PEA} = \Lambda(0.453 - 0.016\,\overline{age} - 0.087\,\overline{educ} - 0.049 \times \overline{vocat}) \cdot (-0.087) \cdot 6$$

$$= -0.522 \times \frac{\exp(0.453 - 0.016\,\overline{age} - 0.087\,\overline{educ} - 0.049 \times \overline{vocat})}{1+\exp(0.453 - 0.016\,\overline{age} - 0.087\,\overline{educ} - 0.049 \times \overline{vocat})}$$

$$\left(\overline{age} = \frac{1}{850}\sum_{i=1}^{850} age_i,\ \overline{educ} = \frac{1}{850}\sum_{i=1}^{850} educ_i,\ \overline{vocat} = \frac{1}{850}\sum_{i=1}^{850} vocat_i \right)$$

(c) $$\widehat{APE} = \frac{1}{850}\sum_{i=1}^{850} \{ \Lambda(0.453 - 0.016\,age_i - 0.087\,educ_i - 0.049 \times 1)$$

$$- \Lambda(0.453 - 0.016\,age_i - 0.087\,educ_i - 0.049 \times 0) \}$$

(d) ATE is same as APE in (c) since vocat is a binary variable & treatment

(e) The effects of age, educ would not be constant and Λ considers this point.

Age term might have diminishing effect, since the increase of age can work well to specific level, but after this level, the increase of age cannot contribute well to decrease the probability of being poverty as people lose their strength, energy, etc. Therefore, the coefficient of age would be negative, while the coefficient of age^2 would be positive

For educ, up to some level, education can decrease the probability of being poverty. But after some level, increase of years of schooling would not dramatically decrease the probability of being poverty. this means education would have diminishing effect with which the coefficient of educ is negative, the coefficient of $educ^2$ is positive. We can plot this like 

The effect of vocat would be constant since it is a binary variable.
We cannot consider the unit increase of binary variable.

2. (a) Let's consider the single linear regression model

$y_i = x_i\beta + \varepsilon_i$, $E(\varepsilon_i | x_i) = 0$

Probit model uses CDF of standard normal distribution to make $Pr(y_i = 1 | x_i)$ lie in $[0, 1]$

Probit model: $Pr(y_i = 1 | x_i) = G(x_i\beta) = \Phi(x_i\beta) = \int_{-\infty}^{x_i\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

In other words, $y_i = \Phi(x_i\beta) + v_i$ is a probit model.

We can estimate β by non-linear least square method:

$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \Phi(x_i\beta))^2$

(b) H_0: The coefficient of mar, $\beta_{mar} = 0$

H_A: The coefficient of mar, $\beta_{mar} \neq 0$

We can do t-test to test the null hypothesis.

By OLS, $t_{mar}^{OLS} = \frac{\hat{\beta}_{mar}^{OLS} - 0}{\sqrt{Var(\hat{\beta}_{mar}^{OLS})}} = \frac{0.024}{0.009} = \frac{8}{3} > 2$

By Logit, $t_{mar}^{Logit} = \frac{\hat{\beta}_{mar}^{Logit} - 0}{\sqrt{Var(\hat{\beta}_{mar}^{Logit})}} = \frac{0.103}{0.057} < 2$

By probit, $t_{mar}^{Probit} = \frac{\hat{\beta}_{mar}^{Probit} - 0}{\sqrt{Var(\hat{\beta}_{mar}^{Probit})}} = \frac{0.063}{0.035} = 1.8 < 2$

We might reject H_0 from OLS, while we might not from logit and probit model.

However, those three models all have heteroskedasticity, leading us not to believe t-statistic in each model.

we should use robust standard errors for three models to get reliable results.

(c) ① OLS : $Pr(y=1 \mid high=1, noqual=0, age=40, age2=16, mar=1)$

$= 0.093\times 1 + (-0.210)\times 0 + 0.038\times 40 + (-0.051)\times 16 + 0.024 - 0.068$

$= 0.753$

② Probit : $Pr(y=1 \mid high=1, noqual=0, age=40, age2=16, mar=1)$

$= \Phi(0.093 + 0.038\times 40 + (-0.051)\times 16 + 0.024 - 0.068)$

$= \Phi(0.753) = 0.774$

The results are similar to each other.
However, if we consider a woman with same condition except aged 80, OLS tells us a negative probability, which makes no sense. In this case, probit can still give us a probability in [0,1]. Therefore, probit estimates are much more proper than OLS if we have to estimate the probability

(d) If we use OLS, the estimated probability can be smaller than 0 or be larger than 1, which is weird. On the other hand, logit and probit models can output the result whose range is in [0,1].
Moreover, OLS assumes that the partial effect of each variable is constant such that it cannot capture varying effect of the variable, while logit and probit models can capture it due to the use of kernel.

Generally, probit can give us an equal or better fit than logit except "extreme independent variables." These are independent variables in which one particularly large or small value will overwhelmingly determine whether the dependent variable is 0 or 1. It overrides the effect of other variables. In this case, logit performs better. But without this scenario, it's better to use probit as long as we can put up with large computation cost.