

# Review questions and exercises: Topic 6

## Before you start

Make sure you are working in the correct directory: `pwd`. Change to another directory if necessary. Also, you may wish to open a command log and a result log:

```
cmdlog using logname.do
log using logname.log
```

## 1 Binary dependent variables

- 1) Use the National Labour Survey data: `sysuse nlsw88, clear`.
- 2) Generate a variable holding the weekly earnings (wage  $\times$  hours).
- 3) Generate a variable that is equal to one if the person is in the top 25% of weekly earners, and zero otherwise. (Hint: use `summarize, detail`)
- 4) Run a regression model where the your new indicator variable is the dependent variable and age, grade, race (factors), married, union, south, smsa, hours, ttl\_exp, and tenure are the independent variables. Use linear regression first. `estimates store` your model.
- 5) Examine the output and comment on sign and significance of the coefficients. Interpret two coefficient estimates of your choice.
- 6) Find out if your linear regression results in predictions outside the (0,1) interval, and how many observations are affected.

- 7) Next, estimate the same model using `logit` . `estimates store` your model.
- 8) Comment on the output. How can you interpret the same two coefficients now?
- 9) Look at goodness of fit statistics for your logit model.
- 10) For your logit model, calculate the average marginal effect of experience, separately for married and single people in your sample. Repeat for the linear regression model, and comment on the differences.
- 11) For your logit model, calculate the *predicted probability* of being a high earner for a white person in the south, who has otherwise average characteristics, at `(grade=(10 13 16))` . Repeat for the linear regression model, and comment.

## 2 Interaction effects

- 1) Expand your model by adding an interaction effect of your choice between a dummy variable (`dvar`) and a continuous variable (`cvar`).
- 2) Estimate by `logit` . `estimates store` your model.
- 3) Calculate the *average marginal effect* of `cvar` on the probability of being a high earner, `over(dvar)` .
- 4) Estimate the same model by OLS. `estimates store` your model.
- 5) Calculate the same marginal effect as in 3). Comment on the difference.
- 6) Choose a second continuous variable (`cvar2`) . `summarize` it to find out its range. For your logit interaction model, calculate the same marginal effect as in 3), `over(dvar) at(cvar2=(a(b)c))` , plugging in sensible values for `a` , `b` and `c` . Plot the marginal effects.
- 7) Repeat 6) for your ols interaction model. Comment on the difference in the graphs.