

Probit and Logit Models Example

Ani Katchova

Probit and Logit Model Example

- We study the factors influencing the purchase of health insurance.
- Using data set from the Health and Retirement Study (HRS), wave 5 (2002) collected by the National Institute of Aging.
- Dependent variable: whether or not a person has health insurance (0 or 1).
- Independent variables: retired, age, good health status, household income, education years, married, Hispanic.
- Estimating regression model, logit, and probit models.

| Health insurance | y codes | Percent frequency |
|------------------|---------|-------------------|
| Yes | 1 | 39% |
| No | 0 | 61% |

Binary outcome model coefficients

| Have health insurance | Regression coefficients | Logit coefficients | Probit coefficients |
|-----------------------|-------------------------|--------------------|---------------------|
| Retired | 0.04* | 0.19* | 0.11* |
| Age | -0.002 | -0.01 | -0.008 |
| Good health status | 0.06* | 0.31* | 0.19* |
| HH income | 0.0004* | 0.002* | 0.001* |
| Education years | 0.02* | 0.11* | 0.07* |
| Married | 0.12* | 0.57* | 0.36* |
| Hispanic | -0.12* | -0.81* | -0.46* |
| Constant | 0.12 | -1.71* | -1.06* |
| R2 | 0.08 | 0.07 | 0.07 |

* Indicates significance at the 5% level.

- Coefficient interpretation: Retired individuals (in comparison to non-retired individuals), individuals with good health status, higher household income, higher education, married are *more likely* to have health insurance, and Hispanic are *less likely* to have health insurance.
- The regression, logit and probit coefficients differ by a scale factor (and therefore we cannot interpret the magnitude of the coefficients).

Binary outcome model marginal effects

| Have health insurance | Regression marginal effects | Logit marginal effects at the mean | Logit average marginal effects | Probit marginal effects at the mean | Probit average marginal effects |
|-----------------------|-----------------------------------|---|---|--|--|
| Retired | 0.04* | 0.04* | 0.04* | 0.04* | 0.04* |
| Age | -0.002 | -0.003 | -0.003 | -0.003 | -0.003 |
| Good health status | 0.06* | 0.07* | 0.06* | 0.07* | 0.06* |
| HH income | 0.0004* | 0.0005* | 0.0005* | 0.0004* | 0.0004* |
| Education years | 0.02* | 0.02* | 0.02* | 0.02* | 0.02* |
| Married | 0.12* | 0.12* | 0.12* | 0.13* | 0.12* |
| Hispanic | -0.12* | -0.16* | -0.16* | -0.16* | -0.15* |

- Marginal effects interpretation: Retired individuals are 4% *more likely* to have insurance (in comparison with those that are not retired). For each additional year in education, individuals are 2% *more likely* to have insurance. Hispanics are 16% *less likely* to have insurance than non-Hispanics.
- Note that unlike the coefficients which are different, the marginal effects are almost identical in the three models.
- The marginal effects at the mean and the average marginal effects are almost identical.
- The signs of the coefficients and marginal effects are the same for the logit and probit models.

- The average of predicted probabilities for having insurance is about 38% which is similar to the actual frequency for having insurance.
- The logit and probit models correctly predict 62% of the values and the rest are misclassified.