Lab 8 - logit and probit models

Julieth Santamaria

March 29, 2019

Based on Bailey (2016)'s Real Econometrics, Chapter 12, Exercise 2.

Background

Public attitudes toward global warming influence the policy response to the issue. The dataset `EnvSurvey.cvs` provides data from a nationally representative survey of the U.S. public that asked multiple questions about the environment and energy. The table below describes the variables:

TABLE 12.8 Variables for Global Warming Data

Variable	Description
Male	Dummy variable = 1 for men
White	Dummy variable = 1 for whites
Education	Education, ranging from 1 for no formal education to 14 for professional/doctorate degree (treat as a continuous variable)
Income	Income, ranging from 1 for household income < $5,000 to 19 for household income > $175,000 (treat as a continuous variable)
Age	Age in years
Party7	Partisan identification, ranging from 1 for strong Republican, 2 for not-so-strong Republican, 3 leans Republican, 4 undecided/independent, 5 leans Democrat, 6 not-so-strong Democrat, 7 strong Democrat

Dataset

We want to estimate whether global warming is real and caused by humans (the dependent variable is `HumanCause`) using as independent variables sex, being white, education, income, age, and partisan identification.

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Application

1. Last lab we estimated the following linear probability models (LPM). The objective was to determine which demographic characteristics helped us predict whether someone thinks global warming has a human cause:

$$HumanCause = \beta_0 + \beta_1 Male + \beta_2 White + \beta_3 Education + \beta_4 IncomeCategory + \beta_5 Age + \beta_6 Age^2 + \beta_7 Party7 + u$$

```
 lpm <- lm(humancause~male + white + educ + incomecat + age + I(age^2)+party7, data=envsvy)

# To calculate y hat (the predicted value of the probability)
    prediction1 <- predict(lpm, interval = "prediction")
```

```
## Warning in predict.lm(lpm, interval = "prediction"): predictions on current data refer to _futur
```

```
    summary(prediction1)
```

```
##       fit               lwr                upr         
##  Min.   :-0.1814   Min.   :-1.04157   Min.   :0.6788  
##  1st Qu.: 0.1835   1st Qu.:-0.67135   1st Qu.:1.0379  
##  Median : 0.3355   Median :-0.51974   Median :1.1904  
##  Mean   : 0.3358   Mean   :-0.51900   Mean   :1.1907  
##  3rd Qu.: 0.4910   3rd Qu.:-0.36372   3rd Qu.:1.3459  
##  Max.   : 0.7708   Max.   :-0.08509   Max.   :1.6266
```

2. Use a probit regression to estimate the probability of saying that global warming is real and caused by humans. Use the independent variables from part (a), including the age-squared variable. Compare statistical significance with LPM results.

```
# Probit
probit <- glm(humancause~male + white + educ + incomecat + age +  I(age*age)+ party7,
              data=envsvy, family=binomial(link="probit"))
summary(probit)
```

```
##
## Call:
## glm(formula = humancause ~ male + white + educ + incomecat +
##     age + I(age * age) + party7, family = binomial(link = "probit"),
##     data = envsvy)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6551  -0.8644  -0.5602   1.0361   2.4695
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.1189279  0.3244321  -6.531 6.52e-11 ***
## male          0.0627495  0.0642795   0.976  0.32897
## white         0.1038861  0.0777799   1.336  0.18167
## educ          0.0810659  0.0185483   4.371 1.24e-05 ***
## incomecat     0.0097614  0.0081337   1.200  0.23009
## age          -0.0286358  0.0104961  -2.728  0.00637 **
## I(age * age)  0.0002569  0.0001055   2.436  0.01486 *
## party7        0.2646544  0.0169363  15.626  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2367.9  on 1854  degrees of freedom
## Residual deviance: 2061.8  on 1847  degrees of freedom
##   (17 observations deleted due to missingness)
## AIC: 2077.8
##
## Number of Fisher Scoring iterations: 4
```

```
# Logit
logit <- glm(humancause~male + white + educ + incomecat + age +  I(age*age)+ party7,
             data=envsvy, family=binomial(link="logit"))
summary(probit)
```

```
##
## Call:
## glm(formula = humancause ~ male + white + educ + incomecat +
##     age + I(age * age) + party7, family = binomial(link = "probit"),
##     data = envsvy)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6551  -0.8644  -0.5602   1.0361   2.4695
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.1189279  0.3244321  -6.531 6.52e-11 ***
## male          0.0627495  0.0642795   0.976  0.32897
## white         0.1038861  0.0777799   1.336  0.18167
## educ          0.0810659  0.0185483   4.371 1.24e-05 ***
## incomecat     0.0097614  0.0081337   1.200  0.23009
## age          -0.0286358  0.0104961  -2.728  0.00637 **
## I(age * age)  0.0002569  0.0001055   2.436  0.01486 *
## party7        0.2646544  0.0169363  15.626  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2367.9  on 1854  degrees of freedom
## Residual deviance: 2061.8  on 1847  degrees of freedom
##   (17 observations deleted due to missingness)
## AIC: 2077.8
##
## Number of Fisher Scoring iterations: 4
```

3. Calculate the marginal effects of being a male. Notice that as male is a dummy variable, it doesn't make sense to change that variable from 0 to 0.4.

Option 1: Marginal Effects at the Means

Being a male increases the probability of support by 0.0213 (about 2.1 percentage points).

```
means <- t(colMeans(envsvy)) # t() makes the vector of means into a 1-row matrix.
means
```

```
##      humancause   party7     white     educ      male     age    agesq
## [1,]         NA 4.144231 0.7617521 12.33066 0.5117521 49.7703 2769.804
##      incomecat treatment warmagree warmagreecontinuous
## [1,]  13.98878  2.476496 0.6928419                  NA
```

```
at.means <- as.data.frame(rbind(means,means))  # 2 copies of the row in a dataframe
at.means$male <- c(0,1)  # female, male
at.means
```

```
##   humancause   party7     white     educ male     age    agesq incomecat
## 1         NA 4.144231 0.7617521 12.33066    0 49.7703 2769.804  13.98878
## 2         NA 4.144231 0.7617521 12.33066    1 49.7703 2769.804  13.98878
##   treatment warmagree warmagreecontinuous
## 1  2.476496 0.6928419                  NA
## 2  2.476496 0.6928419                  NA
```

- Prediction using LPM and MEMs

```
## prediction using LPM
pred.lpms <- predict(lpm, newdata=at.means)
pred.lpms.female <- pred.lpms[1] #prediction on probability for a female
pred.lpms.male <- pred.lpms[2] # prediction on probability for a male
print(pred.lpms.male - pred.lpms.female)
```

```
##          2
## 0.01977601
```

- Prediction using Probit and MEMs

```
## prediction using Probit
pred.probs <- predict(probit, type="response", newdata=at.means)
pred.probs.female <- pred.probs[1] #prediction on probability for a female
pred.probs.male <- pred.probs[2]  #prediction on probability for a male
print(pred.probs.male - pred.probs.female)
```

```
##          2
## 0.02134665
```

Option 2: Average Marginal Effects

Using the AMEs after probit, we find that the effect of going from female to male for all observations is to increase the probability of support by 0.01976 (about 2 percentage points). This is a tiny bit smaller than the LPM estimated effect.

```
pred.envsvy.m <- envsvy
pred.envsvy.m$male <- 1
pred.envsvy.f <- envsvy
pred.envsvy.f$male <- 0
```

- Prediction using LPM and AMEs

```
## prediction using LPM
# prediction on probability for a female
pred.lpm.f <- predict(lpm, newdata=pred.envsvy.f)
# prediction on probability for a male
pred.lpm.m <- predict(lpm, newdata=pred.envsvy.m)
MEs <- pred.lpm.m - pred.lpm.f
print(mean(MEs))          # compute the average of marginal effects
```

```
## [1] 0.01977601
```

- Prediction using Probit and AMEs

```
## prediction using Probit
# prediction on probability for a female
pred.prob.f <- predict(probit, type="response", newdata=pred.envsvy.f)
# prediction on probability for a male
pred.prob.m <- predict(probit, type="response", newdata=pred.envsvy.m)
MEs <- pred.prob.m - pred.prob.f
print(mean(MEs))        # compute the average of marginal effects
```

```
## [1] 0.01976387
```