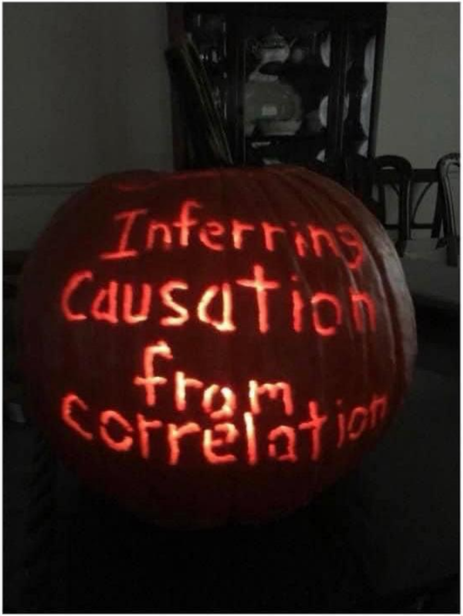
Nothing scares me more...



2017-10-29, 4:03 PM

Recitation 7: Logit and Probit

14.32 Fall 2023

Ian Sapollnik

Nov 3, 2023

What are logit and probit?

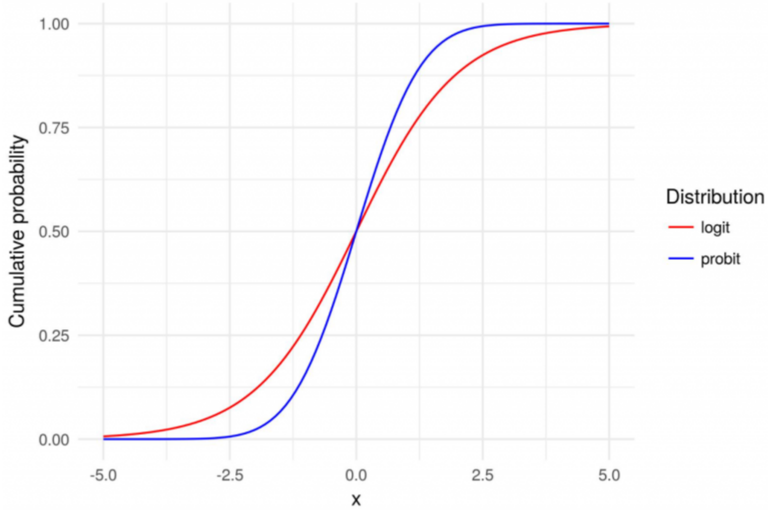
The model we consider is $P\left(Y=1 \mid X\right)=F\left(\alpha+\beta X\right)$, where F is called a *link function*.

A good link function takes us from the real numbers to the interval between 0 and 1: $\mathbb{R} \rightarrow [0,1]$.

We could in theory use any link function, but logit and probit are the most common:

- $F_{\text{logit}}\left(x\right)=\Lambda\left(x\right)=\frac{1}{1+e^{-x}}$
- $F_{\text{probit}}\left(x\right)=\Phi\left(x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$

Logit and probit functions



Picking between logit and probit

- In general there are no hard & fast rules, but rules of thumb
- Look at error terms!
- If many outliers, use logit as it is more robust
- Logit has a cleaner interpretation, as log odds
- When in doubt... common to default to logit
- Discussion a little more complex with dealing with multinomial logit/probit

Interpreting logit and probit models

In general we are interested in marginal effects

$$\frac{\partial P\left(Y=1 \mid X\right)}{\partial x_j} = \frac{\partial F\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i\right)}{\partial x_j} = f\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i\right) \beta_j$$

Interpreting logit and probit models

In general we are interested in marginal effects

$$\frac{\partial P\left(Y=1 \mid X\right)}{\partial x_j} = \frac{\partial F\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i\right)}{\partial x_j} = f\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i\right)\beta_j$$

Fundamental problem: effects we predict depend on the levels of our variables! We saw two approaches in class to deal with this.

Interpreting logit and probit models

In general we are interested in marginal effects

$$\frac{\partial P\left(Y=1 \mid X\right)}{\partial x_j}=\frac{\partial F\left(\beta_0+\sum_{i=1}^{k} \beta_i x_i\right)}{\partial x_j}=f\left(\beta_0+\sum_{i=1}^{k} \beta_i x_i\right) \beta_j$$

Fundamental problem: effects we predict depend on the levels of our variables! We saw two approaches in class to deal with this.

- Partial marginal effect at average

$$\frac{\partial P\left(Y=1 \mid X=\bar{x}\right)}{\partial x_j}=f\left(\beta_0+\sum_{i=1}^{k} \beta_i \bar{x}_i\right) \beta_j$$

Interpreting logit and probit models

In general we are interested in marginal effects

$$\frac{\partial P\left(Y=1 \mid X\right)}{\partial x_j} = \frac{\partial F\left(\beta_0 + \sum_{i=}^{k} \beta_i x_i\right)}{\partial x_j} = f\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i\right)\beta_j$$

Fundamental problem: effects we predict depend on the levels of our variables! We saw two approaches in class to deal with this.

- Average partial effect

$$\frac{1}{n}\sum_{\ell=1}^{n} \frac{\partial P\left(Y=1 \mid X=x_\ell\right)}{\partial x_j} = \frac{1}{n}\sum_{\ell=1}^{n} f\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_{i,\ell}\right)\beta_j$$

Example: predicting NCAA tournament qualification

Follow along with Stata code posted on Canvas if you wish!

Question: do expert rankings predict who can qualify for the NCAA final tournament?

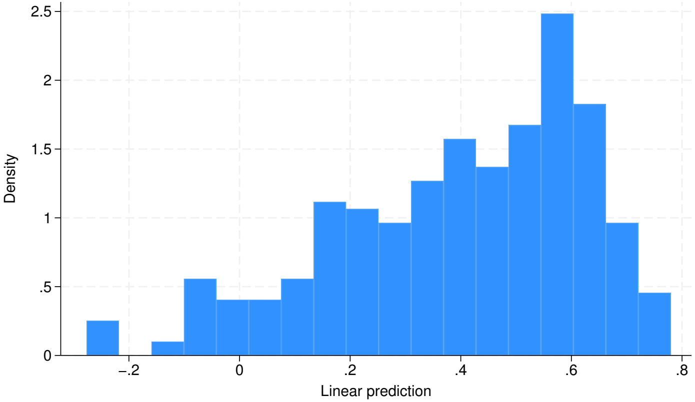
```
. * summary stats
. sum tourney prerpi postrpi_1 coachexper  power5
```

Variable	Obs	Mean	Std. dev.	Min	Max
tourney	336	.3988095	.4903837	0	1
prerpi	336	75.05655	64.12944	1	323
postrpi_1	336	86.59524	72.90892	1	316
coachexper	336	24.94048	8.843382	1	46
power5	336	.7619048	.4265529	0	1

Linear probability model

```
. * trying out a linear probability model
. reg tourney prerpi coachexper  power5, r

Linear regression                               Number of obs     =        336
                                                F(3, 332)         =      38.27
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2066
                                                Root MSE          =     .43877

------------------------------------------------------------------------------
             |               Robust
     tourney | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      prerpi |  -.0033347   .0003496    -9.54   0.000    -.0040224   -.0026469
  coachexper |    .007508   .0028785     2.61   0.010     .0018456    .0131703
      power5 |  -.0562974   .0599212    -0.94   0.348    -.1741706    .0615757
       _cons |   .5047385   .0927727     5.44   0.000     .3222421    .6872349
------------------------------------------------------------------------------

. reg tourney prerpi postrpi_1 coachexper  power5, r

Linear regression                               Number of obs     =        336
                                                F(4, 331)         =      31.55
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2181
                                                Root MSE          =     .43623

------------------------------------------------------------------------------
             |               Robust
     tourney | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      prerpi |   .0000206   .0012818     0.02   0.987    -.0025008    .0025421
   postrpi_1 |  -.0030031   .0011465    -2.62   0.009    -.0052584   -.0007478
  coachexper |   .0069718   .0028751     2.42   0.016     .0013161    .0126276
      power5 |  -.0276566   .0595865    -0.46   0.643    -.1448726    .0895595
       _cons |   .5045044   .0920316     5.48   0.000     .3234639    .6855449
------------------------------------------------------------------------------
```

Predicted values are negative!



Logit

```
. logit tourney prerpi coachexper  power5, r

Iteration 0:  Log pseudolikelihood = -225.96874
Iteration 1:  Log pseudolikelihood = -184.44301
Iteration 2:  Log pseudolikelihood = -182.21799
Iteration 3:  Log pseudolikelihood = -182.19819
Iteration 4:  Log pseudolikelihood = -182.19819

Logistic regression                                     Number of obs =    336
                                                        Wald chi2(3)  =  42.88
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -182.19819                       Pseudo R2     = 0.1937

------------------------------------------------------------------------------
             |               Robust
     tourney | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      prerpi |  -.0224412   .0039588    -5.67   0.000    -.0302003   -.0146821
  coachexper |   .0313247   .0155884     2.01   0.044     .0007721    .0618774
      power5 |  -.1632856    .337554    -0.48   0.629    -.8248793     .498308
       _cons |   .3455622   .5024758     0.69   0.492    -.6392723    1.330397
------------------------------------------------------------------------------
```

```
. logit tourney prerpi postrpi_1 coachexper  power5, r

Iteration 0:  Log pseudolikelihood = -225.96874
Iteration 1:  Log pseudolikelihood = -183.76753
Iteration 2:  Log pseudolikelihood = -181.59773
Iteration 3:  Log pseudolikelihood = -181.56745
Iteration 4:  Log pseudolikelihood = -181.56745

Logistic regression                                     Number of obs =    336
                                                        Wald chi2(4)  =  46.05
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -181.56745                       Pseudo R2     = 0.1965

------------------------------------------------------------------------------
             |               Robust
     tourney | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      prerpi |  -.0103837   .0101899    -1.02   0.308    -.0303555    .0095882
   postrpi_1 |  -.0099673   .0078573    -1.27   0.205    -.0253674    .0054328
  coachexper |   .0308211   .0156325     1.97   0.049     .0001819    .0614603
      power5 |  -.0838392   .3295318    -0.25   0.799    -.7297097    .5620312
       _cons |   .2729752   .4997466     0.55   0.585    -.7065101    1.252461
------------------------------------------------------------------------------
```

What is the predicted probability for LSU?

$$P(\text{tourney} = 1 \mid X) = \Lambda(0.2730 - 0.0104\text{prerpi} - 0.0100\text{postrpi\_1} \\ + 0.0308\text{coachexper} - 0.0838\text{power5})$$

LSU in 2015-16 had a prerpi of 45, postrpi_1 of 65, coachexper of 31 and is in the power5. What is the predicted probability of them reaching the playoffs?

What is the predicted probability for LSU?

$$P\,(\text{tourney} = 1 \mid X) = \Lambda\,(0.2730 - 0.0104\text{prerpi} - 0.0100\text{postrpi\_1} \\ +0.0308\text{coachexper} - 0.0838\text{power5})$$

LSU in 2015-16 had a prerpi of 45, postrpi_1 of 65, coachexper of 31 and is in the power5. What is the predicted probability of them reaching the playoffs?

$$P_{LSU}\,(\text{tourney} = 1) = \Lambda\,(0.2730 - 0.0104 \times 45 - 0.0100 \times 65 \\ +0.0308 \times 31 - 0.0838 \times 1) \\ = \Lambda\,(0.026) \\ = 0.5065$$

How would a more experienced coach affect LSU?

$$P_{LSU}\left(\text{tourney} = 1\right) = \Lambda\left(0.026\right)$$

LSU in 2015-16 had a prerpi of 45, postrpi_1 of 65, coachexper of 31 and is in the power5. What is their marginal effect of a better coach?

How would a more experienced coach affect LSU?

$$P_{LSU}\,(\text{tourney} = 1) = \Lambda\,(0.026)$$

LSU in 2015-16 had a prerpi of 45, postrpi_1 of 65, coachexper of 31 and is in the power5. What is their marginal effect of a better coach?

$$\frac{\partial P\,(Y = 1 \mid X = \text{LSU})}{\partial \text{coachexper}} = \lambda\,(0.026) \times \beta_{\text{coachexper}}$$

$$= 0.25 \times 0.0308 = 0.77\%$$

What is the partial marginal effect of better coaching at the average?

What is the partial marginal effect of better coaching at the average?

$$
\begin{aligned}
\frac{\partial P\left(Y=1 \mid X=\bar{x}\right)}{\partial x_{\text{coachexper}}} =& \lambda\left(\beta_0+\sum_{i=1}^{k} \beta_i \bar{x}_i\right) \beta_{\text{coachexper}} \\
=& \lambda\Big(0.2730-0.0104 \overline{\text{prerpi}}-0.0100 \overline{\text{postrpi\_1}} \\
& +0.0308 \overline{\text{coachexper}}-0.0838 \overline{\text{power5}}\Big) \times 0.0308 \\
=& \lambda\left(0.2730-0.0104 \times 75.06-0.0100 \times 86.60\right. \\
& \left.+0.0308 \times 24.94-0.0838 \times 0.76\right) \times 0.0308 \\
=& \lambda\left(-0.6692\right) \times 0.0308 \\
=& 0.224 \times 0.0308 \\
=& 0.69\%
\end{aligned}
$$

What is the partial marginal effect of better coaching at the average?

```
. margins, dydx(coachexper) atmeans

Conditional marginal effects                          Number of obs = 336
Model VCE: Robust

Expression: Pr(tourney), predict()
dy/dx wrt:  coachexper
At: prerpi     = 75.05655 (mean)
    postrpi_1  = 86.59524 (mean)
    coachexper = 24.94048 (mean)
    power5     = .7619048 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  coachexper |   .0069132   .0035621     1.94   0.052    -.0000684    .0138947
------------------------------------------------------------------------------
```

Addendum: logit and probit are ultimately quite similar