Simple Logistic Regression

Think back to Week 7, when we used the sample from the California Health Indicator Survey (CHIS) to examine the relationship between poverty and visiting the doctor within the past 12 months. This week, we use logistic regression to examine this relationship. Open the `chis_healthdisparities.dta` dataset.

Fit a logistic regression model with visiting the doctor in the past 12 months as the outcome and the poverty indicator as your covariate.

1. List the assumptions for performing logistic regression.

   We assume the responses are Bernoulli, and we assume linearity in the parameters on the logit scale.

2. State your model.

   Define $Y_i = 1$ if individual i visited the doctor in the last 12 months, 0 otherwise. Define $X_i = 1$ if the individual is **above** the poverty line, 0 otherwise. Then, our model is $Y_i \sim Bernoulli(p_i)$, where

$$logit(p_i) = \alpha + \beta X_i$$

3. Fit the model.

```
. logit doctor nopov

Iteration 0:   log likelihood =  -247.4035
Iteration 1:   log likelihood = -245.14765
Iteration 2:   log likelihood = -245.08244
Iteration 3:   log likelihood = -245.08242

Logistic regression                               Number of obs   =        500
                                                  LR chi2(1)      =       4.64
                                                  Prob > chi2     =     0.0312
Log likelihood = -245.08242                       Pseudo R2       =     0.0094

------------------------------------------------------------------------------
      doctor |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       nopov |   .6713351   .3013476     2.23   0.026     .0807047    1.261965
       _cons |     .83975   .2745156     3.06   0.002     .3017093    1.377791
------------------------------------------------------------------------------
```

   The fitted regression model is $logit(\hat{p}_i) = 1.511 + 0.671X_i$.

4. Interpret the coefficients.

- $\alpha =$ log(odds of visiting the doctor when $X_i = 0$)
- $\beta =$ log(odds ratio of visiting the doctor for no poverty versus poverty) = log(odds of visiting doctor when $X_i = 1$) - log(odds of visiting doctor when $X_i = 0$)
- $\alpha + \beta =$ log(odds of visiting the doctor when $X_i = 1$)

5. Provide an OR and a 95% confidence interval.

   Hard way: $exp(\beta) = 1.957$ with 95% CI $(exp(0.0807047), exp(1.261965)) = (1.084, 3.532)$.

   Easy way:

```
. lincom nopov, eform

 ( 1)  [doctor]nopov = 0

------------------------------------------------------------------------------
      doctor |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.956848   .5896914     2.23   0.026     1.084051    3.532357
------------------------------------------------------------------------------
```

   Another easy way:

```
. logistic doctor nopov

Logistic regression                               Number of obs   =        500
                                                  LR chi2(1)      =       4.64
                                                  Prob > chi2     =     0.0312
Log likelihood = -245.08242                       Pseudo R2       =     0.0094

------------------------------------------------------------------------------
      doctor | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       nopov |   1.956848   .5896914     2.23   0.026     1.084051    3.532357
       _cons |   2.315788     .63572     3.06   0.002     1.352168     3.96613
------------------------------------------------------------------------------
```

6. What is the probability of visiting the doctor in the past 12 months for those above poverty? below poverty?

```
. predict phat
(option pr assumed; Pr(doctor))
```

   Below poverty: 0.6984126
   Above poverty: 0.819222

7. Test the hypothesis that $H_0 : \beta = 0$ versus $H_0 : \beta \neq 0$ at the 0.05 level of significance.
   $\hat{\beta} = 0.6713351$, $\hat{se}(\hat{\beta}) = .3013476$, $Z = -2.23$.

   Under H_0, $Z \sim N(0, 1)$, and $p = 0.026$. We reject H_0 and conclude being above the poverty level is associated with higher odds of visiting the doctor within the past 12 months.

Note that the 95% CI for β excludes 0 and the 95% CI for the odds ratio excludes 1, leading to the same conclusion (as will always be the case).

8. Compare your results to the 2×2 table analysis from week 7.

   Yes, our results match up to the contingency table analysis, as they should! The beauty of logistic regression is in its flexibility, as we see next.