Recap + Model Fit

Open the `chis_healthdisparities.dta` dataset.

1. After fitting and considering several models, what are our conclusions about the relationship between poverty and visiting the doctor in the past 12 months?

   Those below the poverty line appear less likely to visit the doctor in the past 12 months.

2. Compare the fit of these models.

   There are several options for assessing the fit of a logistic regression model. We don't have time to look at all of them (if you are interested, look at Hosmer-Lemeshow and deviance). But, to relate back to week 3, let's look at the ROC curve.

   Fit the logistic regression model with `doctor` as the outcome and `nopov`, `female`, and `age` as covariates.

   We choose a cut-off c and construct a classification table:

	$Y_i = 1$	$Y_i = 0$
$\hat{p}_i > c$	Correct	False +
$\hat{p}_i <= c$	False -	Correct

For example, when $c = 0.8$:

```
. estat classification, cutoff(0.8)

Logistic model for doctor

              -------- True --------
Classified |         D             ~D  |      Total
-----------+--------------------------+-----------
     +     |       243             25  |        268
     -     |       159             73  |        232
-----------+--------------------------+-----------
   Total   |       402             98  |        500

Classified + if predicted Pr(D) >= .8
True D defined as doctor != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   60.45%
Specificity                     Pr( -|~D)   74.49%
Positive predictive value       Pr( D| +)   90.67%
Negative predictive value       Pr(~D| -)   31.47%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   25.51%
False - rate for true D         Pr( -| D)   39.55%
False + rate for classified +   Pr(~D| +)    9.33%
False - rate for classified -   Pr( D| -)   68.53%
```

```
--------------------------------------------------
Correctly classified                          63.20%
--------------------------------------------------
```

To get the full ROC curve (and the area under the ROC curve), try `lroc`.

Plot the ROC curve for the three models above to visualize the improved classification of the more complex models. We could likely add more covariates to further improve the discriminatory ability of the model.