Inference for GLMs

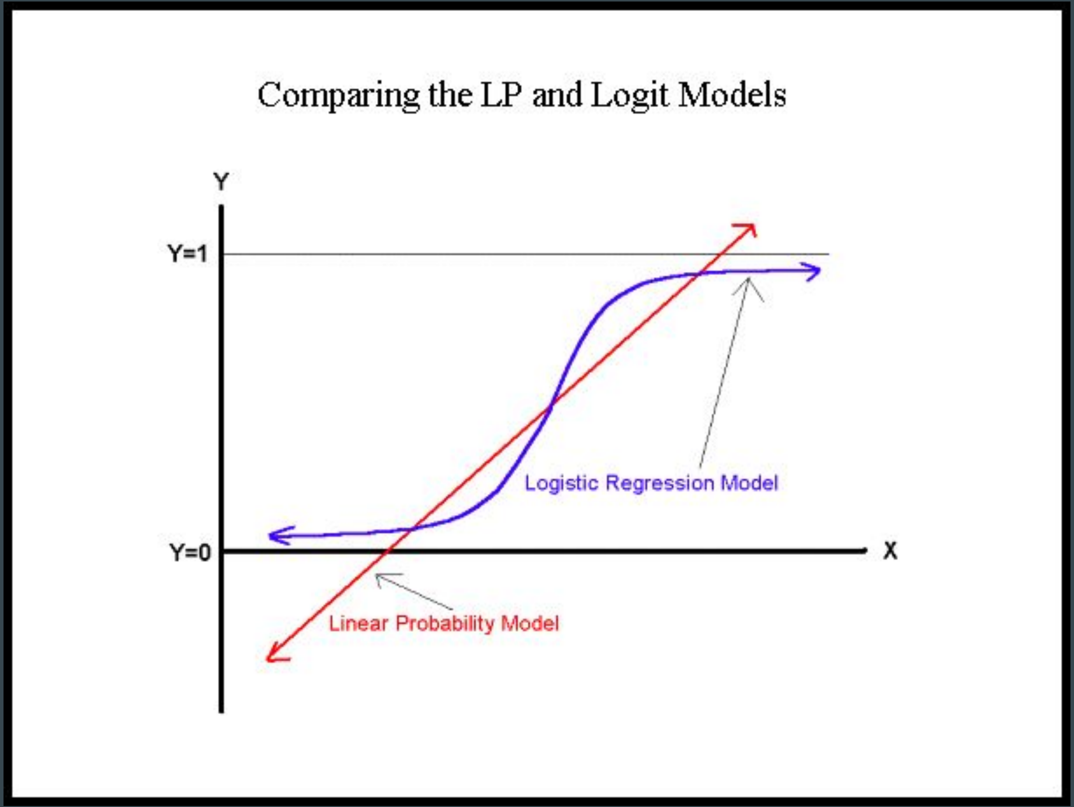
PLSC 309
15 April 2019

Review: GLM = A Linear Model for Non-linear Data

We have seen that a GLM allows us to use a linear framework for non-linear relationships. There are three additional components:

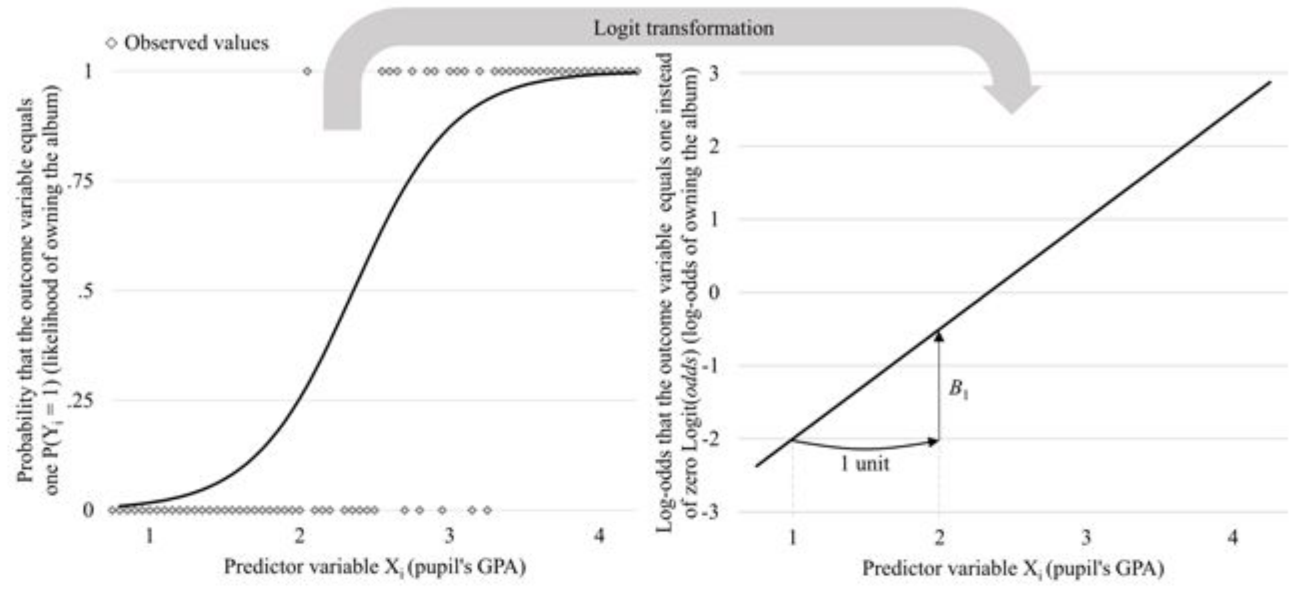
1. Assume a probability distribution (sometimes called the random family)
2. Adjust conditional means (link function) to linear
3. Adjust variance to linear (variance function)

Review: Models for Binary Data



Review: Logit Transformation

$$P(Y_i = 1) = \frac{exp(B_0 + B_1{}^*X_i)}{1 + exp(B_0 + B_1{}^*X_i)}$$



Review: Logistic Regression

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.989979   1.139951  -3.500 0.000465 ***
gre          0.002264   0.001094   2.070 0.038465 *
gpa          0.804038   0.331819   2.423 0.015388 *
rank2       -0.675443   0.316490  -2.134 0.032829 *
```

Benefits of Additivity

- An additive model is one where X variables can be added or subtracted, without changing the β's of other X's
- This means we can decompose the effects of all our explanatory variables and look at the conditional effect of a single X variable on Y
- The usual caveats about interactions and omitted variables apply

Logistic Regression example

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	5.504243	0.504763	10.905	< 2e-16	***
Pclass	-1.161111	0.125412	-9.258	< 2e-16	***
Sexmale	-2.730095	0.200441	-13.620	< 2e-16	***
Age	-0.039934	0.007811	-5.113	3.18e-07	***
SibSp	-0.314241	0.108340	-2.901	0.00373	**
Parch	-0.074519	0.116077	-0.642	0.52089	
EmbarkedQ	-0.048968	0.381622	-0.128	0.89790	
EmbarkedS	-0.453050	0.231819	-1.954	0.05066	.

Logistic Regression example

What is a better predictor of death on the titanic, age or whether or not you have a sibling?

$$\beta_{SIB} = -0.31$$

$$\beta_{AGE} = -0.04$$

You cannot look at the magnitude of β

$$b_1 = \frac{s_y}{s_x} R$$

- Standard deviation is expressed in units of X
- So if X is age (let's say it ranges from 0 to 100), it will *always be larger* than whether you have a sibling (which ranges from 0 to 1)

So how do we tell which variables are important?

This requires a NHST!

$$H_0: \beta = 0$$

$$H_A: \beta \neq 0$$

- No reason for choosing 0, just convention
- Answers whether or not you think it's likely that there is any effect effect of X on Y, or whether their conditional relationship is independent (i.,e. Information about X tells you nothing about Y)

Z-score

We can evaluate our β under a normal distribution, due to the CLT

- For this, we need to get a Z-score (X-axis of a standard normal distribution)
- This is easy: PE / SE
  - PE = point estimate (what you estimated for b)
  - SE = standard error (st. dev divided by N)

Z-score example

$$\beta_{SIB} = -0.31$$

$$\beta_{AGE} = -0.04$$

$$\text{p-value}_{SIB} = -0.31 \,/\, .108 = -2.87 = .002$$

$$\text{p-value}_{AGE} = -0.04 \,/\, 0.008 = -5 = 0.00000029$$

P-value review

The P-value for sibling status is .002, while p-value for age is .00000028. Does this mean that the effect of age is larger?

P-value review

The P-value for sibling status is .002, while p-value for age is .00000028. Does this mean that the effect of age is larger?

NO! It means that we can be more confident in our rejection of the null hypothesis. In other words, we're more convinced that *something is going on* in the case of age, but this has no impact on whether age leads to a greater chance of dying

T-test

- Instead of a z-score, we can also evaluate with a t-distribution
- The test statistic is exactly the same: PE / SE
- Only difference is the t-distribution requires less observations

Why does this work?

- Using the formula β / SE(β)= *t*/*z* is known as a *Wald Test*
- The original Wald Test actually uses χ^2 distribution, but this is almost identical to the normal

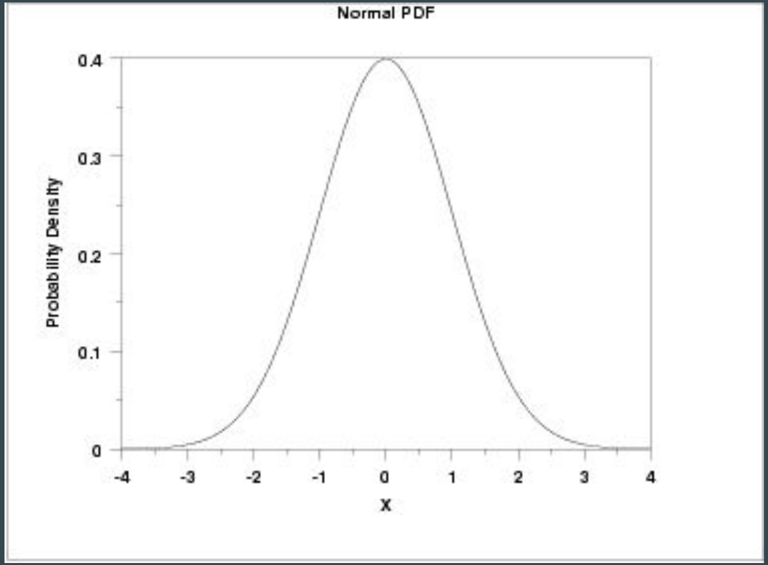
Wald Test

Let θ represent the parameter of a distribution we're trying to estimate (usally θ is β)

$$W = \theta_X - \theta_0 / \text{SE-hat}$$

θ_0 in law of large numbers

If a parameter is actually zero (i.e., there is a slope of 0 in a GLM), in infinite trials, it will be distributed like:



SE-hat

The tricky part of the Wald test is estimating the standard error. To do this, you need to work directly the the log-likelihood derivative.

- We will not be doing this in class
- The lack of symmetry (because we're working with non-normal distributions) will complicate things
- For some distributions, this is impossible to calculate, making the Wald Test impractical

Likelihood Ratio Test (LRT)

- There is a more direct way to test whether or not including a variable is signifciant
- To understand, let's back up
  - The goal of the GLM is to understand P(Y|X)
  - We do this by assuming conditional independence
  - Our p-value, or statistical significance tests, want to find whether $P(Y|X_i)$ are different from 0.
- In other words, does X_i at all increase or decrease the P(Y) conditional on the other X variables?
- MLE gives us the tools to answer this

MLE review

- PDF: probability of y_i for a single observation: x_i
- Likelihood: probability for entire Y vector given entire Y matrix

$$L(p) = \prod_{i=1} p^{x_i} (1-p)^{(1-x_i)}$$

- In other words, the likelihood gives us our conditional probability!

Additivity to the rescue

- The GLM is an additive model
- We know from additivity that we can take out one of our X variables and it shouldn't change the underlying relationships
- This gives us the key to our likelihood ratio

Likelihood Ratio Test

- LRT compares two models:
  - Null model - the model without X_i
  - Alternative model - the model with X_1
- It evaluates whether or not the including your variable increases the likelihood of Y
- In other words, does X_i have any independent effect on Y, holding the other X variables constant?

Likelihood Ratio Test

$$
\begin{aligned}
D &= -2\ln\left(\frac{\text{likelihood for null model}}{\text{likelihood for alternative model}}\right) \\
&= 2\ln\left(\frac{\text{likelihood for alternative model}}{\text{likelihood for null model}}\right) \\
&= 2 \times [\ln(\text{likelihood for alternative model}) - \ln(\text{likelihood for null model})]
\end{aligned}
$$

LRT example

Say we want to build a model to explain whether students got high or low essay scores on a standardized test. We want to predict this using the explanatory variables of gender, math scores, reading comprehension scores, and science scores.

```
Logistic regression                             Number of obs   =        200
                                                LR chi2(4)      =     105.99
                                                Prob > chi2     =     0.0000
Log likelihood = -84.419842                     Pseudo R2       =     0.3857

------------------------------------------------------------------------------
     hiwrite |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.805528   .4358101     4.14   0.000     .9513555      2.6597
        read |   .0529536   .0275925     1.92   0.055    -.0011268     .107034
        math |   .1319787   .0318836     4.14   0.000      .069488     .1944694
     science |   .0577623    .027586     2.09   0.036     .0036947     .1118299
       _cons |  -13.26097   1.893801    -7.00   0.000    -16.97275    -9.549188
------------------------------------------------------------------------------
```

LRT example

- We want to see if there is a meaningful impact of both math and science scores. So we estimate a new model without those X variablesl

```
Logistic regression                               Number of obs   =        200
                                                  LR chi2(2)      =      69.94
                                                  Prob > chi2     =     0.0000
Log likelihood = -102.44518                       Pseudo R2       =     0.2545

------------------------------------------------------------------------------
     hiwrite |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.403022   .3671964     3.82   0.000     .6833301    2.122713
        read |   .1411402   .0224042     6.30   0.000     .0972287    .1850517
       _cons |  -7.798179   1.235685    -6.31   0.000    -10.22008   -5.376281
------------------------------------------------------------------------------
```

LRT example

Test statistic = 2 * (-84 - (-102)) = 36

P-value < .001

Wald vs LRT

- The LRT allows us to:
  - Estimate a p-value even when we can’t calculate standard error
  - Test the combined effect of multiple variables
  - Requires software to compute the log-likelihood
- The Wald Test:
  - Can be computed by hand
  - Does not require re-estimating the model

Review

- We learned two ways of doing inference for regression coefficients
  - Wald Test
  - Likelihood Ratio Test
- However, these are both NHST, and do not tell us anything about the magnitude of our coefficients