Discrete Response Model
Lecture 2

Hypothesis Tests for the Binomial Logistic Regression Parameters

Setting Up the Hypothesis

We often want to assess the importance of an explanatory variable or groups of explanatory variables. One way to make this assessment is through using hypothesis tests. For example, suppose we are interested in the r^{th} explanatory variable x_r in the model

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \quad + \beta_r x_r + \quad + \beta_p x_p$$

If $\beta_r = 0$, we see that x_r would be excluded from the model. Thus, we are interested in hypothesis tests of the form:

Alternatively, we could state the hypotheses as:

$$H_0: \text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \quad + \beta_{r-1} x_{r-1} + \beta_{r+1} x_{r+1} + \quad + \beta_p x_p$$
$$H_a: \text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \quad + \beta_r x_r + \quad + \beta_p x_p$$

- The null hypothesis model terms are all included within the alternative hypothesis model: The null hypothesis model is a special case of the alternative hypothesis model.
- The null hypothesis model is often referred to as a <u>reduced model</u>, and the alternative hypothesis model is often referred to as a <u>full model</u>.

The Wald Test

The Wald statistic

$$Z_0 = \frac{\hat{\beta}_r - \beta_r}{\sqrt{Var(\hat{\beta}_r)}}$$

is used to test $H_0: \beta_r = 0$ vs. $H_a: \beta_r \neq 0$.

For a large sample, the test statistic has an approximate standard normal distribution if the null hypothesis of $\beta_r = 0$ is true. Thus, reject the null hypothesis if a test statistic value is "unusual" for a standard normal distribution, where "unusual" is defined to be $|Z_0| > Z_{1-\alpha/2}$.

The p-value is $2P(Z > |Z_0|)$ where Z ~ N(0,1).

In R, the Wald test statistics and p-values are automatically provided for individual β parameters using code like *summary(mod.fit)*.

The Wald test can also be performed for more than one parameter at the same time. However, because the Wald test has similar problems to those we saw in Lecture 1 for Wald tests and confidence intervals, we are not going to pursue how to perform these types of tests.

Likelihood Ratio Test

Recall that the likelihood ratio test statistic is defined as

$$\Lambda = \frac{\text{Maximum of likelihood function under } H_0}{\text{Maximum of likelihood function under } H_0 \text{ or } H_a}$$

To perform a test of $H_0: \beta_r = 0$ vs. $H_a: \beta_r \neq 0$, we obtain the estimated probabilities of success from estimating

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \quad + \beta_{r-1} x_{r-1} + \beta_{r+1} x_{r+1} + \quad + \beta_p x_p$$

(suppose the estimates are $\hat{\pi}_i^{(0)}$, $\hat{\beta}_0^{(0)}$, … $\hat{\beta}_p^{(0)}$) and the estimated probabilities of success from estimating

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \quad + \beta_r x_r + \quad + \beta_p x_p$$

(suppose the estimates are $\hat{\pi}_i^{(a)}$, $\hat{\beta}_0^{(a)}$, … $\hat{\beta}_p^{(a)}$). We can then find

$$-2\log(\Lambda) = -2\log\left(\frac{L(\hat{\beta}^{(0)} \mid y_1, \ldots, y_n)}{L(\hat{\beta}^{(a)} \mid y_1, \ldots, y_n)}\right) = -2\left[\sum_{i=1}^{n} y_i \log\left(\frac{\hat{\pi}_i^{(0)}}{\hat{\pi}_i^{(a)}}\right) + (1 - y_i)\log\left(\frac{1 - \hat{\pi}_i^{(0)}}{1 - \hat{\pi}_i^{(a)}}\right)\right]$$

If the null hypothesis was true, $-2\log(\Lambda)$ has an approximate χ_1^2 distribution for a large sample.

More General Hypotheses

The hypotheses can be generalized to allow for **q** different βs in H_0 to be set to 0.

Under the null hypothesis, $-2\log(\Lambda)$ has an approximate χ^2_q distribution for a large sample.

Example: Using the Wald Test

Recall from our previous example in which both *change* and *distance* are used.

```
mod.fit2<-glm(formula = good ~ change + distance, family = binomial(link = logit), data =
placekick)
summary(mod.fit2)
```

```
Call:
glm(formula = good ~ change + distance, family = binomial(link = logit),
    data = placekick)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7061   0.2282   0.2282   0.3750   1.5649

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.893181   0.333184  17.687   <2e-16 ***
change      -0.447783   0.193673  -2.312   0.0208 *
distance    -0.112889   0.008444 -13.370   <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: 1013.4  on 1424  degrees of freedom
Residual deviance:  770.5  on 1422  degrees of freedom
AIC: 776.5

Number of Fisher Scoring iterations: 6
```

Below is a summary of the tests using $\alpha = 0.05$.

Change	**Distance**
$H_0: \beta_1 = 0$ $H_a: \beta_1 \neq 0$	$H_0: \beta_2 = 0$ $H_a: \beta_2 \neq 0$
$Z_0 = -2.31$	$Z_0 = -13.37$
p-value = 0.0208	p-value $< 2\times10^{-16}$
Reject H_0 because p-value is small	Reject H_0 because p-value is small
There is sufficient evidence to indicate change has an effect on the probability of success given distance is in the model.	There is sufficient evidence to indicate distance has an effect on the probability of success given change is in the model.

Example: Using the Likelihood Ratio Test

- There are a number of ways to perform LRTs in R.
- The easiest way is to use the **Anova() function** from the **car package**, authored by Professor John Fox.

```
> library(car)
> Anova(mod = mod.fit2, test = "LR")
Analysis of Deviance Table (Type II tests)

Response: good
         LR Chisq Df Pr(>Chisq)
change      5.246  1      0.022 *
distance  218.650  1     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Difference in p-values

- The p-value for distance is the same as from using Anova(), but the p-value for change is not.
- The reason is due to the hypotheses being tested.
  - The Anova() function tests the model's explanatory variables in an sequential manner. Thus, the change test p-value is actually for the test of

$$H_0: \text{logit}(\pi) = \beta_0$$
$$H_a: \text{logit}(\pi) = \beta_0 + \beta_1 \text{change}$$

because it is listed first in the formula argument of glm(). The distance variable is listed second so anova() tests:

$$H_0: \text{logit}(\pi) = \beta_0 + \beta_1 \text{change}$$
$$H_a: \text{logit}(\pi) = \beta_0 + \beta_1 \text{change} + \beta_2 \text{distance}$$

where change is assumed to be in both models.

Produce the Tests Similar to Anova( )

In order to produce the tests like Anova(), estimate the H_0 and H_a models separately and then use their model fit objects in a different way with Anova():

```
> mod.fit2<-glm(formula = good ~ change + distance, family = binomial(link = logit), data = df)
> mod.fit.Ho<-glm(formula = good ~ distance, family = binomial(link = logit), data = df)
> anova(mod.fit.Ho, mod.fit2, test = "Chisq")
Analysis of Deviance Table

Model 1: good ~ distance
Model 2: good ~ change + distance
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1423     775.75
2      1422     770.50  1   5.2455    0.022 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- This use of Anova() helps to emphasize a reduced- and full-model approach to obtaining the $-2\log(\Lambda)$ statistic.
- This approach is helpful to know how to do when Anova() may not be available (with more complex models than logistic regression) or when more than one variable is being tested at a time.

Berkeley SCHOOL OF INFORMATION