Lecture 9

January 31, 2019 9:30 AM

STATISTICS 475: Applied Discrete Data Analysis

Inferences on Probabilities and Odds Ratios

(B&L Section 2.2.3–2.2.4)

1 Problem to be solved

- We know how to do inference on logistic regression model parameters using MLEs $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$.
- We now need to develop inferences for actual probabilities and comparisons of probabilities.
  - I will reverse the order of these topics compared to how they are presented in the book
- *This is where we start to get complicated. Try to keep up, and ask if something doesn't make sense.*
  - Most technical details are given in footnotes. Not essential to learn, but useful for those who like to know why things work.

2 Estimating probabilities from logistic regression

- Estimating a probability of success for a particular value of the explanatory variables involves simply plugging the parameter estimates into the model:

$$\hat{\pi} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p)}$$

plug in MLEs

  - Let's start calling $\text{logit}(\pi) = \log[\pi/(1-\pi)] = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$ the LINEAR PREDICTOR
- Estimate of π is an MLE because each $\hat{\beta}_j$ is a MLE

- MLE's have distributions that we approximate with a normal
- Therefore $\hat{\pi}$ has a distribution that can be approximated by a normal in large samples.
    * In practice, need *very* large samples for this to be close
    * $Var(\hat{\pi})$ can be approximated, but process is complicated and very rough in smaller samples.
- Distribution of $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p$ has a *much* better normal approximation[1]
    * $Var(\hat{\beta} + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p)$ is found directly from the variances and covariances of the regression parameter MLEs[2]
    * As always, better approximation in larger samples

- Therefore, **Wald inferences** are immediately available, by applying them to the linear predictor and converting them into inferences about π

Wald CI works roughly. when n is larger works better.

  - Testing $H_0 : \pi = \pi_0$ at some value of explanatory variables is rarely done
    * If needed, would use Wald Z_w test on $H_0 : \log[\pi/(1-\pi)] = \log[\pi_0/(1-\pi_0)]$, estimating the logit with the estimated linear predictor.

$$Z = \frac{\text{logit}(\hat{\pi}) - \log(\pi_0)}{\sqrt{Var(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}}$$

  - Wald confidence interval for π at any value of explanatory variables is easy to do
    * Find CI for the linear predictor using estimated value and standard error
    * Results in numbers L and U
    * Convert into interval for π using $e^L/(1+e^L)$ and $e^U/(1+e^U)$
        · Reminder that this works on CI, not on standard error, and only works for *monotone* increasing or decreasing functions, not on those than can go both ways
    * Different center and width of interval at each different value of explanatories

$$C.I. = (\hat{\beta}_0 + \cdots + \hat{\beta}_n X_n) \pm Z_{1-\alpha/2}\sqrt{\text{var}\,L} \rightarrow (L, U)$$

$$\Downarrow$$

$$\left(\frac{e^L}{1+e^L}, \frac{e^U}{1+e^U}\right)$$

- Likelihood ratio is preferred but more complicated, as always.
  - LR test for $H_0 : \pi = \pi_0$ at given x_j's requires being able to estimate the model under the constraint $\pi = \pi_0$[3]

[1] Linear combinations of normal random variables are normal random variables

[2] e.g., when $p = 1$ we use the fact that $\widehat{Var}(\hat{\beta}_0 + \hat{\beta}_1 x) = \widehat{Var}(\hat{\beta}_0) + x^2\widehat{Var}(\hat{\beta}_1) + 2x\widehat{Cov}(\hat{\beta}_0, \hat{\beta}_1)$. When $p > 1$ we use the same idea but there are more terms. Stat/Act Sc students should know this.

[3] Equivalent to $\text{logit}(\pi) = \text{logit}(\pi_0)$, so just need to find the profile of the likelihood for which the constraint

$$\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = \text{logit}(\pi_0) \quad \text{(a constant)}$$

holds. For example, when $p = 1$, $\pi = 0.5$ implies $\beta_0 + \beta_1 x = 0$, so just

1. Replace β_0 with $\beta_1 x$ in the likelihood
2. Estimate β_1 as the only parameter remaining in the model
3. Evaluate the resulting log likelihood or residual deviance
4. Compute the test statistic

Of course this is more complicated when $p > 1$, but it can still be done.

– One constraint on parameters under H_0, so χ^2_1

- For (profile) LR confidence interval, need to repeat this process for different potential values of π_0 to find boundaries between those that result in rejecting $H_0 : \pi = \pi_0$ and those that don't.

  – Again, more complicated than Wald, but not impossible.

Example: Placekicking (Lecture 9 scripts.R, Placekick.csv)

We will compute 95% confidence intervals for the probability of successful kick using the model $\text{logit}(\pi) = \beta_0 + \beta_1$`distance`. Recall that we have this fit in an object called `mod.fit`. We show here the profile LR intervals. The programming for the Wald CIs is also in the program for this example.

L.R. in R:

A package called `mcprofile` does a special kind of calculation to quickly approximate the likelihood profile for any linear combination of the regression parameters—that is, function that can be written as

$$a_0\beta_0 + a_1\beta_1 + \ldots + a_p\beta_p.$$

The good news is that most of what we want to compute can be transformed into something of this form. For example, the linear predictor for the probability of successful kick at x yards is just $1 * \beta_0 + x * \beta_1$, so $a_0 = 1$ and $a_1 = x$.

To use the `mcprofile(object=<fitted model>, CM=< "a" matrix>)` function, you need to feed it a model fit object and a matrix of the coefficients (the a's) for each quantity you want a profile for. We will compute LR confidence intervals for the probability of success over the full range of distances in the data set, 18–66 yards.

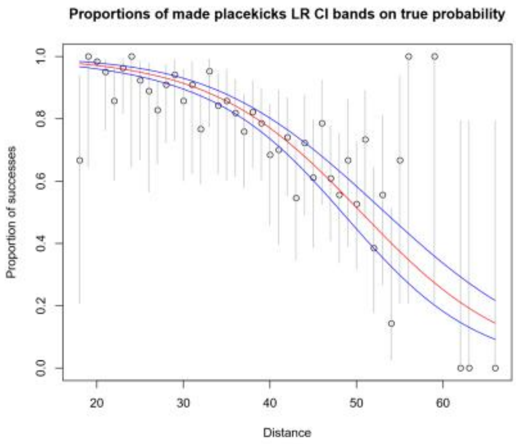
```
> library(package = mcprofile)
>
> # Create the coefficient matrix for the parameters in the
> #   linear predictor
> K <- as.matrix(cbind(1, all.dist))
> class(K)  # matrix
[1] "matrix"
> head(K)
     1 distance
[1,] 1       18
[2,] 1       19
[3,] 1       20
[4,] 1       21
[5,] 1       22
[6,] 1       23
>
> # Use the mcprofile(object=, CM=, ...) function to find profile
> #   likelihood values.
```

```
> # Can take a little time in complex models.
> profiles <- mcprofile(object=mod.fit, CM=K)


> # There is a confint() method developed specifically for objects
> #  produced by mcprofile().  "adjust=" allows you to use methods
> #  to control simultaneous coverage probabilities, like Bonferroni
> ci.logit.LR<-confint(object=profiles, level=0.95, adjust = "none")
> # CI for beta_0 + beta_1 * x
> head(ci.logit.LR$confint) #confint() object contains >1 element
     lower    upper
1 3.389490 4.125899
2 3.288023 3.996258
3 3.186336 3.866807
4 3.084403 3.737572
5 2.982191 3.608581
6 2.879664 3.479865
> ci.pi.LR<-exp(ci.logit.LR$confint)/(1 + exp(ci.logit.LR$confint))
> head(ci.pi.LR)
      lower     upper
1 0.9673745 0.9841077
2 0.9640156 0.9819476
3 0.9603168 0.9795038
4 0.9562448 0.9767420
5 0.9517631 0.9736243
6 0.9468319 0.9701094
```

The results of all of the these confidence intervals are given in Figure 1. Notice that the blue bands and the gray bars are both confidence intervals for the same quantity: the probability of success at each distance. Except in one case, the model-based CIs are *much* narrower than the individual CIs. Why do you think this is?

Figure 1: Placekick data proportions of observed successes with Wilson score confidence intervals at each distance (gray vertical lines), the logistic regression fit (red curve), and the pointwise 95% profile LR confidence limits (blue curves).



3 Estimating odds ratios to compare probabilities

- We often want to estimate the *effect* of a variable on the mean response
  - Like in linear regression: how much does $E(Y)$ change as x_j increases by 1 unit, holding other variables constant
  - "slope", β_j
- We want to do the same thing with logistic regression
  - We might like to estimate how much π (the binary mean) changes as x_j increases, holding other variables constant.
  - This turns out to be not so easy, because of the structure of the model
    * Slope is not constant across range of π
  - What *is* easy is to estimate how much the *odds* of success change.
  - So this is what we will do.
- Logistic regression models the log odds, $\log[\pi/(1-\pi)] = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$
  - Odds of success at a given $x_1, \ldots, x_p$ are just $\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)$
  - Odds ratios (ORs) are just ratios of odds, computed at different levels of $x_1, \ldots, x_p$ in the numerator and denominator
    * $\exp(a)/\exp(b) = \exp(a-b)$
    * So odds ratios are just exponentiated differences between two linear predictors!
    * EASY to compute
- Consider $p = 1$ for illustration, but the following works for any p assuming that you focus only on one x_j and hold others constant
  - Model is $\text{logit}(\pi) = \beta_0 + \beta_1 x_1$
  - The odds of success at $x_1 = x$ are $\text{Odds}_x = \exp(\beta_0 + \beta_1 x)$
  - The odds of success at $x_1 = x+1$ are $\text{Odds}_{x+1} = \exp(\beta_0 + \beta_1(x+1)) = \exp(\beta_1) * \text{Odds}_x$, regardless of the value of x
  - Therefore the change in odds as x increases by 1 unit is measured by the odds ratio

$$OR = \frac{\text{Odds}_{x+1}}{\text{Odds}_x} = \exp([\beta_0 + \beta_1(x+1)] - [\beta_0 + \beta_1 x]) = \exp(\beta_1)$$

    * For a c-unit increase in x_1, the OR is $OR = \exp(c\beta_1)$
  - The odds of success change *multiplicatively* by $\exp(c\beta_1)$ for each c-unit increase in x

- Of course, MLE for OR just substitutes estimated parameters:

$$\widehat{OR} = \exp(c\hat{\beta}_1)$$

- **Inference on OR is easy: Do inferences based on β_1 and exponentiate!**
  - Test for $H_0 : OR = a$ is rarely done, except for $a = 1$, in which case, this is just $H_0 : \beta_1 = 0$.
  - Confidence intervals are found for β_1 using previous methods (Wald or LR) and endpoints are exponentiated
- Changing more than one variable at a time presents no challenge
  - Suppose I have two sets of values for $x_1 \ldots, x_p$, say, $a_1, \ldots, a_p$ and $b_1, \ldots, b_p$
  - $OR = \exp([\beta_0 + \beta_1 a_1 + \ldots + \beta_p a_p] - [\beta_0 + \beta_1 b_1 + \ldots + \beta_p b_p])$
  - Not often done except in special cases that we will see later

Example: Placekicking (Lecture 9 scripts.R, Placekick.csv)

How much easier might a field goal be if the team could get the ball 10 yards closer? We answer this by estimating the OR associated with a 10-yard *decrease* in distance and finding the corresponding 95% LR confidence interval.

Since our model is $\text{logit}(\pi) = \beta_0 + \beta_1$`distance`, this OR is simply

$$OR = \exp(-10\beta_1).$$

Thus, we first need to extract β_1 from `mod.fit`, multiply by -10, and exponentiate. To find the confidence interval, we find a confidence interval for β_1 and repeat the process: multiply the endpoints by -10, and exponentiate.

```
> # Estimated odds ratio
> # 1-yard increase
> exp(mod.fit$coefficients[2])
 distance
0.8913424
> # 10-yard decrease
> exp(-10*mod.fit$coefficients[2])
distance
3.159035
> # Profile likelihood interval for regression parameters
> beta.ci <- confint(object=mod.fit, parm="distance", level=0.95)
Waiting for profiling to be done...
> beta.ci  # C.I. for beta
      2.5 %      97.5 %
-0.13181435 -0.09907103
```

reverse

π

```
> rev(exp(-10*beta.ci))     # Ignore limit labels
   97.5 %     2.5 %
2.693147 3.736478
> as.numeric(rev(exp(-10*beta.ci)))  # Limit labels removed
[1] 2.693147 3.736478
```

So the odds of successful placekick increase by 3.2 times for each 10 yards closer that the kick is taken, with a 95% LR CI from 2.7 to 3.7.

---

4 Notes

1. We can get inferences on the probability of failure just by taking $1-\pi$ without redoing any other calculations
   (a) Hypotheses about $1-\pi$, change direction for 1-sided tests
   (b) Apply directly to confidence interval endpoints
2. *USEFUL CONVERSIONS:* Odds ratios can also be re-stated in a number of ways
   (a) Original calculation is OR =Odds of **Success** in Group **1** / Odds of **Success** in Group **2**
   (b) Odds of **Success** in Group **2** / Odds of **Success** in Group **1** $= 1/OR$
   (c) Odds of **Failure** in Group **1** / Odds of **Failure** in Group **2** $= 1/OR$
   (d) Odds of **Failure** in Group **2** / Odds of **Failure** in Group **1**$= OR$
3. In summary: DEFINE YOUR PARAMETERS CLEARLY so that you know what your analysis is telling you.

5 What to learn from this

1. Here, and in everything else we do, inferences are usually based on quantities for which it is easy to compute sampling distributions
   (a) Linear predictor is easy to work with
   (b) Probabilities and ORs are complicated, but are monotone functions of the linear predictor.
   (c) Wald inferences are easy, but computer can handle LR, which are better. Score is even more difficult than LR and is not usually done
2. Get *very* used to working with comparisons of parameters, like in the OR calculations outlined in Section 3. We will do this again and again, except with more complexity!

6 Exercises (due when announced)

Compete the following exercises from B&L, Chapter 2. As always, add proper interpretations on all confidence intervals and tests.

1. Exercise 5 (c), (d). In (c) instead of Wald, use LR to compute bands. Also, include Wilson Score confidence intervals for each point on the plot, as in Figure 1 above.

2. The longest field goal ever attempted in the NFL is from 76 yards. Using our placekick example above, estimate the probability of a successful placekick from 76 yards, the longest NFL attempt ever made.

   (a) Compute the LR confidence interval for this distance and interpret the results. In particular, address how likely it appears to be that such a long attempt would succeed.

   (b) The longest successful field goal is from 64 yards. By how much are the odds of better from this distance than from 76 yards? Include a confidence interval with tour response.

3. Refer to the model fit to all explanatory variables in the placekick data from Lecture 8.

   (a) Two kickers are arguing at a pub about their statistics. Steve plays for a team that plays most of its games indoors on artificial turf, while Nick's team plays most of its games outdoors on grass. Nick believes that his job is clearly harder than Steve's (he is less likely to be successful, accounting for any other factors) and wants you to support his claim. Use the model to estimate a measure of how much Nick's playing conditions affect his success compared to Steve. Do your results support Nick's claim? Explain.

4. Exercise 14.

In addition, here are exercises that will not be marked, but that you can do for practice:

1. Exercise 15: Stat students should be able to use the delta method to find variances for quantities that are *not* linear functions of the parameters, such as $Var(\hat{\pi})$