# Discrete Response Model
# Lecture 2

datascience@berkeley

# Odds Ratios

# Interpretation (cont.)

There are a number of ways to interpret the odds ratio in the context of logistic regression. The one following is commonly used.

The odds of a success change by $\boxed{e^{c\beta_1} \text{ times}}$ for every c-unit increase in x

If x is a binary (or indicator) explanatory variable having only two levels coded as 0 or 1, then

$$\text{Odds}_{x=0} = e^{\beta_0 + \beta_1 0} = e^{\beta_0} \quad \text{and} \quad \text{Odds}_{x=1} = e^{\beta_0 + \beta_1}$$

$$OR = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

The odds of a success are $e^{\beta_1}$ times as large for x = 1 than for x = 0

# Estimated Odds Ratio

To find the estimated odds ratio, simply replace the parameter with its corresponding estimate:

$$OR = e^{c\hat{\beta}_1}$$

- The interpretation of the odds ratio now needs to be qualified with an "estimated" in the appropriate location in the sentence.
- This estimate is the MLE.

# Wald Confidence Interval for OR

Wald confidence intervals are the easiest to calculate. First, an interval for $c\beta_1$ needs to be found:

$$c\hat{\beta}_1 \pm cZ_{1-\alpha/2}\sqrt{\mathrm{Var}(\hat{\beta}_1)}$$

where $\mathrm{Var}(\hat{\beta}_1)$ is obtained from the estimated covariance matrix for the parameter estimates. Notice where c is located in the interval calculation. The second c comes about through $\mathrm{Var}(c\hat{\beta}_1) = c^2\mathrm{Var}(\hat{\beta}_1)$.

To find the $(1 - \alpha)$ Wald confidence interval for OR, use the exponential function:

$$e^{c\hat{\beta}_1 \pm cZ_{1-\alpha/2}\sqrt{\mathrm{Var}(\hat{\beta}_1)}}$$

# Profile Likelihood Ratio Confidence Interval

- As you might expect, Wald confidence intervals do not always work as well as we would like for smaller sample sizes.
- Instead, a better interval is a <u>profile likelihood ratio interval</u>.
- For a $(1-\alpha)100\%$ interval, we find the set of $\beta_1$ values such that

$$-2\log\left(\frac{L(\beta_0,\beta_1 \mid y_1,\ldots,y_n)}{L(\hat{\beta}_0,\hat{\beta}_1 \mid y_1,\ldots,y_n)}\right) < \chi^2_{1,1-\alpha}$$

  is satisfied.
- On the left side, we have the usual $-2\log(\Lambda)$ form, but without a specified value of $\beta_1$.
- The $\beta_0$ is an estimate of $\beta_0$ for a fixed value of $\beta_1$. Iterative numerical procedures can be used to find the $\beta_1$ values that satisfy the above equation.
- The $(1-\alpha)100\%$ interval for OR is then

$$e^{c\times lower} < OR < e^{c\times upper}$$

  using the lower and upper limits found for $\beta_1$ in the above equation.

# Some Remarks

1) Inverting odds ratios less than 1 is helpful for interpretation purposes.
2) An appropriate value of c should be chosen in the context of the explanatory variable. For example, if $0.1 < x < 0.2$, a value of $c = 1$ would not be appropriate. Additionally, if $0 < x < 1000$, a value of $c = 1$ may not be appropriate as well.
3) When there is more than one explanatory variable, the same interpretation of the odds ratio *generally* can be made with the addition of "holding the other explanatory variables constant" added.
4) If the specification includes transformations of the explanatory variables, the odds ratio is not simply $e^{c\beta r}$ as given previously, because the odds ratio is no longer constant for every c-unit increase in x.
5) A categorical explanatory variable represented by multiple indicator variables does not have the same type of interpretation as given previously.

# Example

Consider the model with only distance as the explanatory variable:

$$logit(\hat{\pi}) = 5.8121 - 0.150 distance$$

To estimate the odds ratio, we can simply use the exp() function:

```
> exp(mod.fit$coefficients[2])
 distance
0.8913424
> 1/exp(10*mod.fit$coefficients[2])
distance
3.159035
```

The first odds ratio is for a 1-yard ($c = 1$) increase in distance. This is not very meaningful in the current context. Instead, $c = 10$ would be more meaningful because 10 yards are needed for a first down in football.

Also, it is more meaningful to look at a 10 yard decrease (another first down) rather than a 10 yard increase. Therefore, the estimated odds of a success change by $\frac{1}{e^{10(-0.1150)}} = 3.16$ times for every 10 yard decrease in the distance of the placekick.

- Note that the 3.16 holds for comparing 30- to 20-yard placekicks as well as 55- to 45-yard placekicks or any other 10-yard decrease.

# Example

To account for the variability in the odds ratio estimator, we would like to calculate a confidence interval for the actual odds ratio itself. Below is the code for the profile likelihood ratio interval:

```
> beta.ci<-confint(object = mod.fit, parm = "distance", level = 0.95)
Waiting for profiling to be done...
> beta.ci
      2.5 %      97.5 %
-0.13181435 -0.09907103
> rev(1/exp(beta.ci*10)) #Invert OR C.I. and c=10
  97.5 %    2.5 %
2.693147 3.736478
> as.numeric(rev(1/exp(beta.ci*10)))
[1] 2.693147 3.736478
```

- The **confint() function** first finds an interval for $\beta_1$ itself.
- We then use the **exp() function** to find the confidence interval for OR.
- The 95% profile likelihood ratio confidence interval is $2.69 < OR < 3.74$.
- Pay special attention to how this was found with the rev() function and the **beta.ci** object.
- Because the interval is entirely above 1, there is sufficient evidence that a 10-yard decrease in distance increases the odds of a successful placekick.

Berkeley SCHOOL OF INFORMATION