Discrete Response Model
Lecture 3

datascience@berkeley

Variable Transformation, Part 1: Interactions Among Explanatory Variables

Odds Ratio

- With interaction effect incorporated in the model, we need to find the odds ratio for wind comparing windy (1) vs. non-windy (0), holding distance constant.
- Recall the formula from our earlier discussion:

$$OR = \frac{Odds_{x_2+c}}{Odds_{x_2}} = \frac{e^{\beta_0+\beta_1 x_1+\beta_2(x_2+c)+\beta_3 x_1(x_2+c)}}{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}} = e^{c(\beta_2+\beta_3 x_1)}$$

- For this equation, x_1 is distance and x_2 is wind. Of course, c = 1 for this setting due to wind being binary.
- Because distance could be anywhere from 18 to 66 yards, we use distances of 20, 30, 40, 50, and 60 yards to interpret the odds ratio for wind.

Interpretation of the Distance Odds Ratio

- For the distance odds ratio, we need to hold wind constant at 0 or 1.

- We also need to choose a value for **c** with respect to distance. The OR equation is

$$OR = \frac{Odds_{x_1+c}}{Odds_{x_1}} = \frac{e^{\beta_0+\beta_1(x_1+c)+\beta_2 x_2+\beta_3(x_1+c)x_2}}{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}} = e^{c(\beta_1+\beta_3 x_2)}$$

```
> round(data.frame(wind = wind, OR.hat = 1/OR.dist, OR.low
+     = 1/exp(ci.log.OR.up), OR.up = 1/exp(ci.log.OR.low)),2)
  wind OR.hat OR.low OR.up
1    0   3.01   2.54  3.56
2    1   6.96   3.03 15.98
```

Notice the odds ratios are inverted. Below are the interpretations:

- With 95% confidence, the odds of a success change by an amount between 2.54 to 3.56 times for every 10-yard decrease in distance under non-windy conditions.
- With 95% confidence, the odds of a success change by an amount between 3.03 to 15.98 times for every 10-yard decrease in distance under windy conditions.

Odds Ratio

```r
# Odds Ratios
beta.hat<-mod.fit.Ha$coefficients[2:4]
c<-1
distance<-seq(from = 20, to = 60, by = 10)

OR.wind<-exp(c*(beta.hat[2] + beta.hat[3]*distance))
cov.mat<-vcov(mod.fit.Ha)[2:4,2:4]

#Var(beta^_2 + distance*beta^_3)
var.log.OR<-cov.mat[2,2] + distance^2*cov.mat[3,3] + 2*distance*cov.mat[2,3]

ci.log.OR.low<-c*(beta.hat[2] + beta.hat[3]*distance) - c*qnorm(p =
0.975)*sqrt(var.log.OR)
ci.log.OR.up<-c*(beta.hat[2] + beta.hat[3]*distance) + c*qnorm(p =
0.975)*sqrt(var.log.OR)

round(data.frame(distance = distance, OR.hat = 1/OR.wind, OR.low =
1/exp(ci.log.OR.up), OR.up = 1/exp(ci.log.OR.low)),2)
```

	distance	OR.hat	OR.low	OR.up
1	20	0.45	2.34	2.34
2	30	1.04	2.71	2.71
3	40	2.41	5.08	5.08
4	50	5.57	20.06	20.06
5	60	12.86	99.13	99.13

Berkeley SCHOOL OF INFORMATION