**Section 3.3.1 – Odds ratios**

A logistic regression model can be written as



Thus, we were modeling the “log odds of a success.” One could also state the odds portion as “odds of a success vs. a failure,” but this was unnecessary because we had a binary response (we could only have a success or failure).

A multinomial regression model can be written as

log(πj/π1) = βj0 + βj1x1 + … + βjpxp for j = 2, …, J

Thus, we are modeling the “log odds of a category j response vs. a category 1 response.” This allows us to use odds ratios to interpret an explanatory variable's relationship with the response variable!

Consider the model again of

 log(πj/π1) = βj0 + βj1x for j = 2, …, J

The odds of a category j response vs. a category 1 response are

πj/π1 = exp(βj0 + βj1x)

The odds ratio for a c-unit increase in x is



Thus, the odds of a category j vs. a category 1 response are  times as large for a c-unit increase in x. Similarly, we could also say the odds of a category j vs. a category 1 response change by



times for every c-unit increase in x.

In a similar manner, we could also compare category j to j′(j ≠ j′, j > 1, j′ > 1) with an odds ratio:



This comes from re-expressing the model as

log(πj/πj′) = βj0 – βj′0 + x(βj1 – βj′1)

which we did previously for j = 2 and j′ = 3.

Notes:

* The odds ratio interpretation specifically states “the odds of a category j vs. a category 1” comparison. In the past when Y was a binary response, we said something like “the odds of a success” only, because it was assumed that a comparison was being made to the one other response category (failure).
* When there is more than one explanatory variable, we will need to include a statement like “holding the other variables in the model constant”.
* Similar to what we saw in Chapter 2, adjustments need to be made for an odds ratio calculation and interpretation when interactions or transformations are present in the model.
* Inference methods for odds ratios are performed in the same ways as for likelihood procedures discussed in earlier chapters.

Example: Wheat kernels (Wheat.R, Wheat.csv)

The class variable is binary, so I need to use a value of c = 1 for it. For the remaining explanatory variables, I will use a value of c equal to 1 standard deviation. Ideally, it would be best to speak with the subject-matter researcher about possible values for c. Below is how I estimate the odds ratios for each explanatory variable.

> sd.wheat <- apply(X = wheat[,-c(1,7)], MARGIN = 2, FUN = sd)

> c.value <- c(1, sd.wheat) #class = 1 is first value

> round(c.value,2)

 density hardness size weight moisture

 1.00 0.13 27.36 0.49 7.92 2.03

> # beta.hat\_jr for r = 1, ..., 6 and j = 2, 3

> # mod.fit$coefficients does not exist, so use coefficients()

> coefficients(mod.fit)

 (Intercept) classsrw density hardness

Scab 30.54650 -0.6481277 -21.59715 -0.01590741

Sprout 19.16857 -0.2247384 -15.11667 -0.02102047

 size weight moisture

Scab 1.0691139 -0.2896482 0.10956505

Sprout 0.8756135 -0.0473169 -0.04299695

> beta.hat2 <- coefficients(mod.fit)[1,2:7]

> beta.hat3 <- coefficients(mod.fit)[2,2:7]

> # Odds ratios for j = 2 vs. j = 1 (scab vs. healthy)

> round(exp(c.value\*beta.hat2), 2)

 density hardness size weight moisture

 0.52 0.06 0.65 1.69 0.10 1.25

> round(1/exp(c.value\*beta.hat2), 2)

 density hardness size weight moisture

 1.91 17.04 1.55 0.59 9.90 0.80

> # Odds ratios for j = 3 vs. j = 1 (sprout vs. healthy)

> round(exp(c.value\*beta.hat3), 2)

 density hardness size weight moisture

 0.80 0.14 0.56 1.54 0.69 0.92

> round(1/exp(c.value\*beta.hat3), 2)

 density hardness size weight moisture

 1.25 7.28 1.78 0.65 1.45 1.09

Example interpretations include:

* The estimated odds of a scab vs. a healthy response are 0.06 times as large for a 0.13 increase in the density, holding the other variables constant. Equivalently, we can say that the estimated odds of a scab vs. a healthy response are 17.04 times as large for a 0.13 decrease in the density, holding the other variables constant.
* The estimated odds of a sprout vs. a healthy response are 7.28 times as large for a 0.13 decrease in the density, holding the other variables constant.
* The estimated odds of a scab vs. healthy response are 9.90 times as large for a 7.92 decrease in the weight, holding the other variables constant.
* The estimated odds of a sprout vs. healthy response are 1.45 times as large for a 7.92 decrease in the weight holding the other variables constant. Note that a Wald test of H0: β35 = 0 vs. Ha: β35 ≠ 0, which uses the regression parameter needed for this sprout vs. healthy odds ratio, had a p-value of 0.2, so this odds ratio may not be interpreted in actual applications.

Relate these odds ratios again back to what we saw in the parallel coordinates plot.

Question: How could estimated odds ratios for sprout vs. scab be calculated? We do not have estimates of regression parameters calculated directly for this comparison!

The confint() function calculates Wald intervals for the regression parameters. Once these intervals are found, one can use the exp() function with these intervals to find intervals for the odds ratios.

> conf.beta <- confint(object = mod.fit, level = 0.95)

> conf.beta # Results are in a 3D array

, , Scab

 2.5 % 97.5 %

(Intercept) 22.14 38.95

classsrw -1.95 0.65

density -27.70 -15.49

hardness -0.04 0.00

size -0.44 2.58

weight -0.41 -0.17

moisture -0.19 0.41

, , Sprout

 2.5 % 97.5 %

(Intercept) 11.78 26.55

<OUTPUT EDITED>

> ci.OR2 <- exp(c.value\*conf.beta[2:7,1:2,1])

> ci.OR3 <- exp(c.value\*conf.beta[2:7,1:2,2])

> round(data.frame(low = ci.OR2[,1], up = ci.OR2[,2]), 2)

 low up

classsrw 0.14 1.92

density 0.03 0.13

hardness 0.37 1.12

size 0.80 3.55

weight 0.04 0.26

moisture 0.67 2.32

> round(data.frame(low = 1/ci.OR2[,2], up = 1/ci.OR2[,1]),

 2)[c(2,5),]

 low up

density 7.64 38.00

weight 3.80 25.79

> round(data.frame(low = ci.OR3[,1], up = ci.OR3[,2]), 2)

 low up

classsrw 0.30 2.13

density 0.07 0.28

hardness 0.36 0.87

size 0.91 2.59

weight 0.39 1.22

moisture 0.58 1.44

> round(data.frame(low = 1/ci.OR3[,2], up = 1/ci.OR3[,1]),

 2)[c(2,3),]

 low up

density 3.57 14.82

hardness 1.15 2.74

The density odds ratios can be interpreted as:

With 95% confidence, the odds of a scab instead of a healthy response are 7.64 to 38.00 times as large when density is decreased by 0.13, holding the other variables constant. Also, with 95% confidence, the odds of a sprout instead of a healthy response are 3.57 and 14.82 times as large when density is decreased by 0.13, holding the other variables constant.

For the scab vs. healthy comparison, only the density and weight odds ratio confidence intervals do not include 1. For the sprout vs. healthy comparison, only the density and hardness odds ratio confidence intervals do not include 1.

What about sprout vs. scab? My recommendation for this case is to use the factor() function to change the order of the levels, this will lead to multinom() estimating one regression parameter for the comparison. The confint() and exp() function can be used again with the new model fit to find the Wald interval.

The calculation process can be simplified (?) by using emmeans. Below are the calculations for Wald confidence intervals involving density:

> calc.est <- emmeans(object = mod.fit, specs = ~ type + density, mode = "latent", at = list(density = c(1.25, 1.25+0.1313021))

> test.info <- contrast(object = calc.est, interaction = list("revpairwise", "pairwise"))

> confint(object = test.info, df = Inf, type = "response", level = 0.95)

 type\_revpairwise density\_pairwise ratio SE df

 Scab / Healthy 1.25 / 1.3813021 17.043 6.973 Inf

 Sprout / Healthy 1.25 / 1.3813021 7.278 2.642 Inf

 Sprout / Scab 1.25 / 1.3813021 0.427 0.104 Inf

 asymp.LCL asymp.UCL

 7.643 38.004

 3.573 14.824

 0.265 0.688

Degrees-of-freedom method: user-specified

Confidence level used: 0.95

Intervals are back-transformed from the log scale

Comments:

* Odds ratios can be calculated for only one explanatory variable at a time.
* The specs argument contains the response variable and the explanatory variable of interest.
* The mode value is latent. This has emmeans() work with a form of the linear predictor.
* Only the density values are needed in the at argument. I use 1.25 and 1.25 + standard deviation (c-unit increase) here, but some other value than 1.25 could be used.
* The interaction argument of contrast() allows us to specify the ordering of the comparisons. The use of list specifies a reverse pairwise comparison for type (e.g., Scab is in a numerator of the odds ratio when compared to Healthy) and a pairwise comparison for density. Try different values of pairwise and revpairwise in list() to see the effect.