W271 Summer 2022 Lecture Video Question Solutions Week 4

Contents

Week 4 Discrete Response Model Part 4

4.4 Independence

Q: Without consulting any resources, write down in mathematical form the definition of independence of two random variable with PDFs P(X) and P(Y).

Solution: There are a few ways we can mathematically described independence of two random variables. Fundamentally though, independence means that the distribution of the two random variables do not depend on one another.

Mathematically this can be represented as:

$P(X = x|Y = y) = P(X = x)$ **and** $P(Y = y|X = x) = P(Y = y)$ **or** $P(X = xandY = y) = P(X = x)P(Y = y)$

4.8 Odds Ratio

Q: Show $e^{c(\beta_{j1}-\beta_{j'1})}$ for a multinomial logistic regression model

Solution: Remember that in multinomial logistic regression, we estimate essentially separate regressions for each category relative to the reference category. The coefficients and odds ratios we get from a multinomial logistic regression are therefore relative to the reference category.

Following, the properties of odds ratios, we multiply and cancel things out to get the desired odds ratios comparing different groups in the model.

Let's say we have the setup from the lecture such that:

$log(\frac{\pi_j}{\pi_1}) = \beta_{j0} + \beta_{j1}x$

$log(\frac{\pi_{j'}}{\pi_1}) = \beta_{j'0} + \beta_{j'1}x$

Therefore, for group j the odds ratio of a c change in x relative to the reference category is $e^{c\beta_{j1}}$.

For group j' the odds ratio of a c change in x relative to the reference category is $e^{c\beta_{j'1}}$, so the odds of the reference group to group j' is $e^{-c\beta_{j'1}}$.

Therefore, the odds ratio of group j to group j' is $\frac{Odds_j}{Odds_{ref}} * \frac{Odds_{ref}}{Odds_{j'}} = \frac{Odds_j}{Odds_{j'}} = e^{c(\beta_{j1}-\beta_{j'1})}$**.**

Q: How could the estimated odds ratios for sprout vs. scab be calculated?

Solution: Following the above prompt, we can exponentiate the difference between coefficients times the desired value to get at the odds ratio for sprout vs. scab.

Say we have the model below:

```
library(nnet)

#data can be found here: https://www.chrisbilder.com/categorical/programs_and_data.html
dat <- read.csv("~/Documents/Berkeley W271/Week 4 Discrete Response Model Part 4/wheat.csv")
head(dat)
```

```
##   class  density hardness    size  weight moisture    type
## 1   hrw 1.349253 60.32952 2.30274 24.6480 12.01538 Healthy
## 2   hrw 1.287440 56.08972 2.72573 33.2985 12.17396 Healthy
## 3   hrw 1.233985 43.98743 2.51246 31.7580 11.87949 Healthy
## 4   hrw 1.336534 53.81704 2.27164 32.7060 12.11407 Healthy
## 5   hrw 1.259040 44.39327 2.35478 26.0700 12.06487 Healthy
## 6   hrw 1.300258 48.12066 2.49132 33.2985 12.18577 Healthy
```

```
multi.model <- multinom(type ~ class + density + hardness + size + weight + moisture,
                        data = dat)
```

```
## # weights:  24 (14 variable)
## initial  value 302.118379
## iter  10 value 234.991271
## iter  20 value 192.127549
## final  value 192.112352
## converged
```

```
summary(multi.model)
```

```
## Call:
## multinom(formula = type ~ class + density + hardness + size +
##     weight + moisture, data = dat)
##
## Coefficients:
##        (Intercept)   classsrw   density    hardness      size     weight
## Scab      30.54650 -0.6481277 -21.59715 -0.01590741 1.0691139 -0.2896482
## Sprout    19.16857 -0.2247384 -15.11667 -0.02102047 0.8756135 -0.0473169
##           moisture
## Scab    0.10956505
## Sprout -0.04299695
##
## Std. Errors:
##        (Intercept)  classsrw  density    hardness      size     weight
## Scab      4.289865 0.6630948 3.116174 0.010274587 0.7722862 0.06170252
## Sprout    3.767214 0.5009199 2.764306 0.008105748 0.5409317 0.03697493
##         moisture
## Scab   0.1548407
## Sprout 0.1127188
##
```

```
## Residual Deviance: 384.2247
## AIC: 412.2247
```

Let's use the class variable. Class is a categorical variable with values of hrw and srw. Therefore, the coefficient in the fitted model below represents the marginal impact of being in the srw class on the logit wher ethe reference category is Healthy.

We can find the odds ratio of sprout to scab for srw by subtracting the coefficients of class from the model for sprout and scab and exponentiating the difference.

```
class.coef <- coefficients(multi.model)[,"classsrw"]
sprout.to.scab.or <- unname(exp(class.coef["Sprout"] - class.coef["Scab"]))
sprout.to.scab.or
```

```
## [1] 1.527129
```

The odds ratio of sprout to scab is 1.5 for class srw, meaning that wheat with a class of srw are 1.5 times more likely to be sprout than scab.