Lecture 12: Gelman Hill Ch 5

Logistic Regression Coefficients

Recall, the (inverse) logistic function will map the output of $X\beta$ to the support of the parameter, p.

add drawing here

At a specified value of the predictor values, we can map the values to a probability. However, the change in one unit for a predictor value does not result in a constant change for the probability.

Typically it is recommended to standardize the variables for interpretation.

GH present a divide by 4 rule that gives a quick, approximate interpretation. THat is that the upper bound for the % difference in the probability for a 1 unit change can be acheived by dividing the value by 4.

The common interpretation of coefficients for logistic regression uses odds and log odds. Let an event have a probability of occurrence $= p$.

Then the odds of that event can be written as: $\frac{p}{1-p}$

- $p = .5$ implies odds of 1 (or 1 to 1)
- $p = 1/4$ implies odds of $\frac{1/4}{3/4} = 1/3$ or "one-to-three"
- $p = 1/10$ implies odds of $\frac{1}{9}$

Note that gambling terminology tends to use odds against your outcome, whereas, the above definitions are odds for. So the odds of having a birthday party on May 15 (without social distancing might be) $9 - to - 1$, which is equal to $\frac{1}{9}$ odds for and implies a .1 probability of occurence.

An *odds ratio* is the ratio of two odds. Say

$$\frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

Log odds

Recall, in logistic regression,

$$log\left(\frac{Pr(y=1|x)}{1-Pr(y=1|x)}\right) = log\left(\frac{Pr(y=1|x)}{Pr(y=0|x)}\right) = \beta_0 + \beta_1 x$$

the term on the left side of the equation is the log odds.

Hard to interpret, but useful for very complicated models and when only direction of relationship is needed.

Odds scale

The effects can also be interpreted on the odds scale by exponentiating the coefficients: $exp(\hat{\beta}_1)$

Probability scale

GH suggest that they prefer to interpret the results on the probability scale.

As we have seen this presents some challenges, especially when considering other parameters (need to be held constant). Results are nonlinear so not easily summarized as either linear or multiplicative changes but can be useful if predicted probabilities are of interest

This might be best shown visually by making a plot of predicted probabilities as a function of the predictor variable

Logistic regression with continuous predictors

1) Log-odds scale: Two subjects that are 1 unit apart in x are predicted to have difference in mean log-odds of improvement of β_1 log-odds (95% CI: ...)

2) Odds scale: Two subjects that are 1 unit apart in x are predicted to have mean odds of improvement that are $\exp(\beta_1)$ times different (95% CI: ...)

3) Probability scale: Needs to specify the start / end values. For instance, when the variable is one SD greater than the mean the probability changes by z % or changes from p_1 to p_2

Logistic regression with single categorical predictor

- The "1 unit" difference in x is now the change to the given category versus the baseline on either the log-odds, odds, or probability scales:

- $logit(\pi) = \beta_0 + \beta_1 I_{Level=2}$

- $\hat{\beta}_0$ is the estimated (mean) log-odds of success in the baseline category

- $exp(\hat{\beta}_0)$ is the estimated (mean) odds of success in the baseline category

- $\hat{\beta}_1$ is the estimated (mean) difference in log-odds of success in "Level = 2" compared to the baseline.

- $exp(\hat{\beta}_1)$ is the estimated (mean) number of times the odds of success is higher or lower for "Level = 2" compared to the baseline.

  - It is also the estimate of the "Odds Ratio", which is the estimated odds of success in given group divided by the odds of success in the baseline group

  - Estimated Odds ratio: $\frac{\hat{\pi}_{L2}/(1-\hat{\pi}_{L2})}{\hat{\pi}_{L1}/(1-\hat{\pi}_{L1})}$

Inference:

- Wald tests: $z = \hat{\beta}_k / SE(\hat{\beta}_k) \sim N(0,1)$

- Wald confidence intervals: $\hat{\beta}_k \mp z^*_{1-\alpha/2} SE(\hat{\beta}_k)$

- Requires asymptotics and delta method

Logistic Regression with interactions/more complex models

- When considering models with multiple predictors or predictors with more than 2 levels, the tools for linear models directly apply on the link scale - for similar subjects on X1, X2, etc., the difference in the estimated mean log-odds of success is $\hat{\beta}_k$ for subjects that are 1 unit different in X_k or in a given category vs reference "WHEN HOLDING ALL OTHER PREDICTORS CONSTANT"

- With interactions, we can get a different slope for a predictor of interest at each level of the other predictor

Diagnostics

- Remember that GLMs have independence assumptions just like linear models

- Residuals for binary logistic models can be rough because of the discreteness in the responses

  - $Residual_i = y_i - \hat{\pi}_i$

- Binned residuals:

3

– Using the `binnedplot` from `arm`, we can find the average $Residual_i$ for groups of predicted values to look for systematic missed curvature

```r
suppressMessages(library(arm))
par(mfrow=c(1,2))
binnedplot(x=predict(glm_int,type="response"),y=as.numeric(Arthritis$Imp2)-1 - predict(glm_int,type="re
binnedplot(x=predict(glm_add,type="response"),y=as.numeric(Arthritis$Imp2)-1 - predict(glm_add,type="re
```