STA 235 - Interpreting Logistic Regression

Spring 2021

McCombs School of Business, UT Austin

Logistic Regression

- Last class we reviewed **logistic regression**. But...

You might still be confused

Do not dispare!

Do not dispare!

More examples coming your way

Logistic Regression

- As we discussed, a logistic regression looks like this:

$$logit(p) = \log(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \varepsilon$$

Where:

- $p = Pr(Y = 1)$ is the probability of "success" (or that Y=1)
- $\log(\frac{p}{1-p})$ is a **log odds** (i.e. the logarithm of an odd, which is the probability of success over the probability of failure).

Logistic Regression

Now, let's imagine two scenarios:

- **Scenario 1**: $X_1 = x_1$, $X_2 = x_2$, ... , and $X_p = x_p$

Then, if plug in these values of $X's$ into our estimated model, we get an expected value of:

$$\log(\frac{p_1}{1-p_1}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p$$

Where p_1 is the expected probability of success given those values of X's.

Now, our **scenario 2** will be *exactly the same*, except that I will increase one unit of X_1. Then,

$$\log(\frac{p_2}{1-p_2}) = \hat{\beta}_0 + \hat{\beta}_1 (x_1 + 1) + \ldots + \hat{\beta}_p x_p$$

Logistic Regression

If we subtract scenario 2 from scenario 1, we get the following:

$$\log(\frac{p_2}{1-p_2}) - \log(\frac{p_1}{1-p_1}) = \hat{\beta}_0 + \hat{\beta}_1(x_1+1) + \ldots + \hat{\beta}_p x_p - (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p)$$

$$\log(\frac{p_2}{1-p_2}) - \log(\frac{p_1}{1-p_1}) = \hat{\beta}_1$$

$$\log(\frac{\frac{p_2}{1-p_2}}{\frac{p_1}{1-p_1}}) = \hat{\beta}_1$$

Which is a **LOG ODDS RATIO!**.

This means that $\hat{\beta}_1$ represents the change in log odds ratio when I increase one unit of X_1, holding other variables constant.

Log Odds Ratios

- Log odds ratios (and odd ratios) are hard to interpret... but let's try.
- Remember we left here:

$$\log\left(\frac{\frac{p_2}{1-p_2}}{\frac{p_1}{1-p_1}}\right) = \hat{\beta}_1$$

- If we exponentiate both sides, we get the following:

$$\exp \hat{\beta}_1 = \frac{\frac{p_2}{1-p_2}}{\frac{p_1}{1-p_1}}$$

which is now an odds ratio... a little bit better.

Odds Ratios

- Remember from last class, that an odds ratio is the odds of something happening in scenario 1 over the odds of something happening in scenario 2.
- An example:
    - "The odds of getting admitted into grad school are 1.5 times higher if you are male applicant than a female applicant."
- What does this means?

$$\frac{Pr(Admitted|Male)}{Pr(NotAdmitted|Male)} = 1.5\frac{Pr(Admitted|Female)}{Pr(NotAdmitted|Female)}$$

- This is the same as saying that the odds of getting admitted into grad school are 50% higher if you are male than you are female.
- *Remember that being twice as likely, means that you are 100% more likely... which is weird.*

Probabilities

- A more intuitive way of looking at this is **estimating probabilities**.
- However, because we are not estimating a linear model, the change in probabilities depends on where we stand in the distribution, and **depends on the values we choose for our other X's**.
- How do we do this?
    - Choose some informative values for your other covariates (you can choose a group of interest, evaluate the variables in their mean/mode, etc.)
    - Plug in your values in your estimated model and calculate the probabilities for each scenario of X_1 and $X_1 + 1$
    - Take the difference! (All of this is included on the R script on the course website)

One more example

Let's look at an additional example: Getting into grad school

```r
d <- read.csv("https://stats.idre.ucla.edu/stat/data/binary.csv")

head(d)
```

```
##   admit gre  gpa rank
## 1     0 380 3.61    3
## 2     1 660 3.67    3
## 3     1 800 4.00    1
## 4     1 640 3.19    4
## 5     0 520 2.93    4
## 6     1 760 3.00    2
```

One more example

Let's run a simple logit model: Being admitted as a function of GPA and GRE score

```r
logit1 <- glm(admit ~ gre + gpa, data = d,  family = binomial(link = "logit"))

logit1
```

```
##
## Call:  glm(formula = admit ~ gre + gpa, family = binomial(link = "logit"),
##     data = d)
##
## Coefficients:
## (Intercept)          gre          gpa
##   -4.949378     0.002691     0.754687
##
## Degrees of Freedom: 399 Total (i.e. Null);  397 Residual
## Null Deviance:       500
## Residual Deviance: 480.3     AIC: 486.3
```

How do we interpret the GPA coefficient?

One more example

Let's run a simple logit model: Being admitted as a function of GPA and GRE score

```r
logit1 <- glm(admit ~ gre + gpa, data = d,  family = binomial(link = "logit"))

logit1
```

```
##
## Call:  glm(formula = admit ~ gre + gpa, family = binomial(link = "logit"),
##     data = d)
##
## Coefficients:
## (Intercept)          gre          gpa
##   -4.949378     0.002691     0.754687
##
## Degrees of Freedom: 399 Total (i.e. Null);  397 Residual
## Null Deviance:       500
## Residual Deviance: 480.3     AIC: 486.3
```

One more point of GPA is associated with a 0.75 increase in log odds of being admitted, holding GRE constant.

One more example

Let's run a simple logit model: Being admitted as a function of GPA and GRE score

```r
logit1 <- glm(admit ~ gre + gpa, data = d,  family = binomial(link = "logit"))

logit1
```

```
##
## Call:  glm(formula = admit ~ gre + gpa, family = binomial(link = "logit"),
##     data = d)
##
## Coefficients:
## (Intercept)          gre          gpa
##   -4.949378     0.002691     0.754687
##
## Degrees of Freedom: 399 Total (i.e. Null);  397 Residual
## Null Deviance:       500
## Residual Deviance: 480.3     AIC: 486.3
```

One more point of GPA is associated with a 0.75 increase in log odds of being admitted, holding GRE constant.

What does that mean??

One more example

Let's run a simple logit model: Being admitted as a function of GPA and GRE score

```r
logit1 <- glm(admit ~ gre + gpa, data = d,  family = binomial(link = "logit"))

logit1
```

```
##
## Call:  glm(formula = admit ~ gre + gpa, family = binomial(link = "logit"),
##     data = d)
##
## Coefficients:
## (Intercept)          gre          gpa
##   -4.949378     0.002691     0.754687
##
## Degrees of Freedom: 399 Total (i.e. Null);  397 Residual
## Null Deviance:        500
## Residual Deviance: 480.3     AIC: 486.3
```

The odds of being admitted into grad school increase by a 2.1 factor with one additional point of GPA (exp(0.75) = 2.1), holding GRE constant.

One more example

Let's run a simple logit model: Being admitted as a function of GPA and GRE score

```
logit1 <- glm(admit ~ gre + gpa, data = d,  family = binomial(link = "logit"))

predict(logit1, newdata = data.frame("gre" = c(mean(d$gre),mean(d$gre)),
                                     "gpa" = c(mean(d$gpa)-0.5,mean(d$gpa)+0.5)),
        type = "response")
```

```
##         1         2 
## 0.2337791 0.3935517
```

The probability of being admitted into grad school increases from 23% to 39% if I increase my GPA from 2.9 to 3.9, holding GRE constant at 588.

One more example

One last note: **All the hypothesis testing are valid with log odds ratio!** (if something is statistically significant in the output, is significant too as change in probabilities, etc.)

```
summary(logit1)
```

```
##
## Call:
## glm(formula = admit ~ gre + gpa, family = binomial(link = "logit"),
##     data = d)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2730  -0.8988  -0.7206   1.3013   2.0620
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.949378   1.075093  -4.604 4.15e-06 ***
## gre          0.002691   0.001057   2.544   0.0109 *
## gpa          0.754687   0.319586   2.361   0.0182 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 480.34  on 397  degrees of freedom
## AIC: 486.34
##
## Number of Fisher Scoring iterations: 4
```