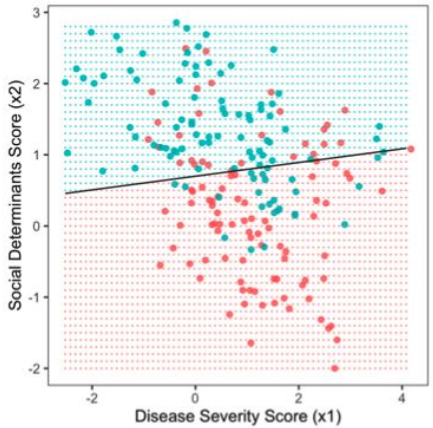
Chapter 9: Interpreting a Logistic Regression Model

Modern Clinical Data Science
Chapter Guides
Bethany Percha, Instructor

How to Use this Guide

- Read the corresponding notes chapter first
- Try to answer the discussion questions on your own
- Listen to the chapter guide (should be 15 min, max) while following along in the notes

$$\log \frac{\mu}{1-\mu} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$




```
Call:
glm(formula = y ~ x1 + x2, family = "binomial", data = df)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.88232  -0.90614  -0.05965   0.86579   2.28489

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.9780     0.2945   3.321 0.000897 ***
x1            0.1344     0.1372   0.980 0.327272
x2           -1.3981     0.2316  -6.035 1.59e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 277.26  on 199  degrees of freedom
Residual deviance: 209.54  on 197  degrees of freedom
AIC: 215.54

Number of Fisher Scoring iterations: 4
```

Question 9.1

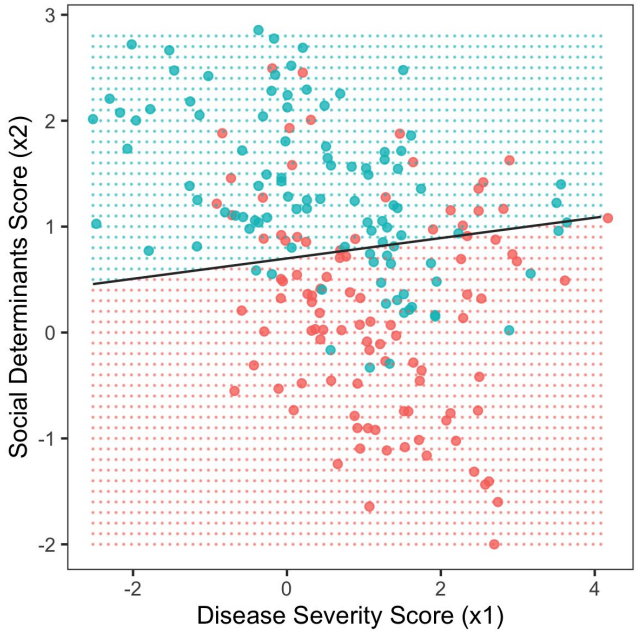
In logistic regression, μ itself is not equal the sum of the predictors; instead, the **logit** of μ is their sum. Based on what you know about μ, why is a logistic regression model not of the form

$$\mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p?$$

$$\log \frac{\mu}{1-\mu} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Question 9.2

The decision boundary in logistic regression (see picture above) occurs where the sum of the linear predictors, $\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$, is zero. What value of μ does this correspond to? Why does this make sense, intuitively?



$$\log \frac{\mu}{1-\mu} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Question 9.3

Looking at the form of the logistic regression model

$$\log \frac{\mu}{1-\mu} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

what does the value of each of the βs mean? What is β_j telling us about how y varies with the predictor j, all else being equal?

Question 9.4

The **odds** of something happening are defined as $\mu/(1-\mu)$, where μ is the probability that the thing occurs. In our example model, we are interested in the odds that $y = 1$ (the patient is readmitted). Does a unit increase in x_1 (disease severity score) increase or decrease the odds that a patient will be readmitted? What about x_2 (social determinants score)?

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.9780     0.2945   3.321 0.000897 ***
x1            0.1344     0.1372   0.980 0.327272
x2           -1.3981     0.2316  -6.035 1.59e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 277.26  on 199  degrees of freedom
Residual deviance: 209.54  on 197  degrees of freedom
AIC: 215.54

Number of Fisher Scoring iterations: 4
```

Question 9.5

What are the odds of readmission for a patient with:

(a) $x_1 = 0.1$ and $x_2 = 0.3$?

(b) $x_1 = 0.1$ and $x_2 = -1.3$?

(c) $x_1 = 1.1$ and $x_2 = 0.3$?

```
Call:
glm(formula = y ~ x1 + x2, family = "binomial", data = df)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.88232  -0.90614  -0.05965   0.86579   2.28489

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.9780     0.2945   3.321 0.000897 ***
x1            0.1344     0.1372   0.980 0.327272
x2           -1.3981     0.2316  -6.035 1.59e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 277.26  on 199  degrees of freedom
Residual deviance: 209.54  on 197  degrees of freedom
AIC: 215.54

Number of Fisher Scoring iterations: 4
```

$$\log\left(\frac{\mu}{1-\mu}\right) = 0.9780 + 0.1344(0.1) - 1.3981(0.3) = 0.572$$

$$\frac{\mu}{1-\mu} = \exp(0.572) = 1.772$$

$$\log\left(\frac{\mu}{1-\mu}\right) = 0.9780 + 0.1344(0.1) - 1.3981(-1.3) = 2.809$$

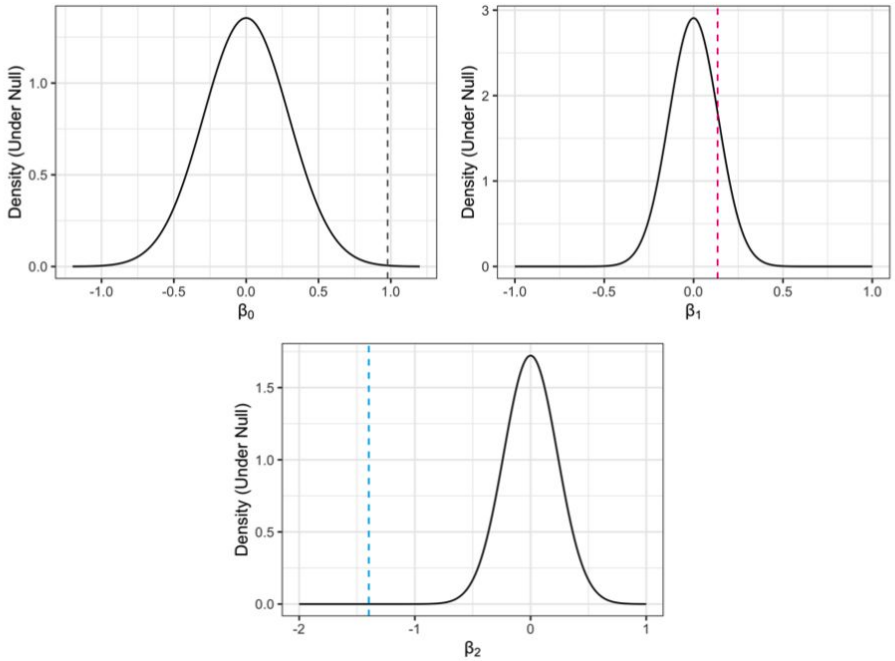
$$\frac{\mu}{1-\mu} = \exp(2.809) = 16.592$$

$$\log\left(\frac{\mu}{1-\mu}\right) = 0.9780 + 0.1344(1.1) - 1.3981(0.3) = 0.706$$

$$\frac{\mu}{1-\mu} = \exp(0.706) = 2.027$$

Question 9.6

Below are the null distributions for the hypothesis tests of our three regression coefficients, β_0, β_1, and β_2. In each graph, the maximum likelihood estimate of the coefficient is shown as a vertical dashed line. Based on these graphs, can you tell why the p-values for β_0 and β_2 are low and the one for β_1 is high? What is the intuition behind this?



Question 9.7

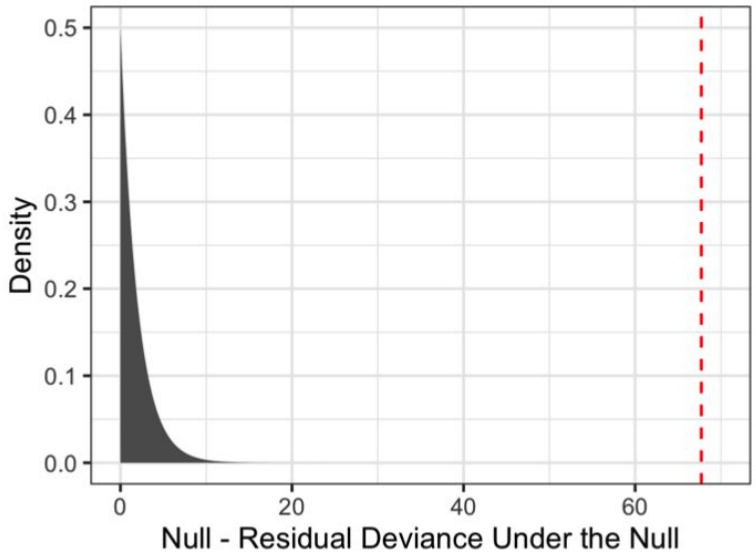
This test is a hypothesis test of the null hypothesis that a model with no predictors fits our data as well as our model, where goodness of fit is measured by the deviance (lower is better). What is this hypothesis test akin to in the linear regression model output?

```
Call:
glm(formula = y ~ x1 + x2, family = "binomial", data = df)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.88232  -0.90614  -0.05965   0.86579   2.28489

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.9780     0.2945   3.321 0.000897 ***
x1            0.1344     0.1372   0.980 0.327272
x2           -1.3981     0.2316  -6.035 1.59e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 277.26  on 199  degrees of freedom
Residual deviance: 209.54  on 197  degrees of freedom
AIC: 215.54

Number of Fisher Scoring iterations: 4
```

```
Call:
lm(formula = y ~ x1 + x2, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-11.9218  -3.1032   0.2891   2.8316  12.5813

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  49.8600     0.5370  92.844  < 2e-16 ***
x1           10.4372     0.2855  36.555  < 2e-16 ***
x2           -1.8824     0.3609  -5.215 4.63e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.769 on 197 degrees of freedom
Multiple R-squared:  0.9026,    Adjusted R-squared:  0.9016
F-statistic: 912.4 on 2 and 197 DF,  p-value: < 2.2e-16
```

Question 9.8

The difference in null and residual deviances in this case is 67.72. It follows a χ^2_2 distribution under the null. A plot of the χ^2_2 distribution and our test statistic is shown below. What do these findings indicate about the p-value of this goodness of fit test and what does it mean?



```
Call:
glm(formula = LOW ~ AGE + LWT + RACE + SMOKE + PTL + HT + UI +
    FTV, family = "binomial", data = d)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8946  -0.8212  -0.5316   0.9818   2.2125

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.480623   1.196888   0.402  0.68801
AGE         -0.029549   0.037031  -0.798  0.42489
LWT         -0.015424   0.006919  -2.229  0.02580 *
RACE2        1.272260   0.527357   2.413  0.01584 *
RACE3        0.880496   0.440778   1.998  0.04576 *
SMOKE        0.938846   0.402147   2.335  0.01957 *
PTL          0.543337   0.345403   1.573  0.11571
HT           1.863303   0.697533   2.671  0.00756 **
UI           0.767648   0.459318   1.671  0.09467 .
FTV          0.065302   0.172394   0.379  0.70484
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 201.28  on 179  degrees of freedom
AIC: 221.28

Number of Fisher Scoring iterations: 4
```

Question 9.9

In this model, is the effect of one predictor (say, `AGE`) impacted by the value(s) of any of the other predictor(s)? How does this differ from the other classification algorithms we've seen (KNN and decision trees)? What are the advantages and disadvantages of this choice?

Question 9.10

Comment on how the variable `RACE` enters into the model here. Does this make sense in light of what that variable means and how it potentially interacts with the other study variables?

Question 9.11

Interpret the values of each of these coefficients. Based on the coefficient values and their standard errors, which predictor(s) do you think have the greatest impact on whether or not a woman has a low birthweight baby?