# 5.1: Logistic Regression

Dr. Bean - Stat 5100

## 1 Why Logistic Regression?

Recall the example from 5.1.1 where we try to predict the odds of having a disease using:

- Age (in years)
- Sector (1 if living in high risk city sector, 0 otherwise)

**Based on the output in Figure 1 what is the *probability* that a 40 year old living in the high risk city sector has the disease?**

$$\hat{L} = -2.335 + 0.0293 * 40 + 1.6734 * 1 = 0.5104$$

$$\hat{\pi} = \frac{1}{1 + e^{-0.5104}} = 0.6249$$

There is a 62.49% chance.

**Using the same information provided in Figure 1, provide an interpretation of the coefficient associated with Age?**

Holding city sector constant, a one year increase in age multiplies the odds of having the disease by $e^{0.0293} = 1.0297$

Holding city sector constant, the odds of having the disease are $100(e^{0.0293} - 1) = 2.97\%$ greater for each year increase in Age.

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| **Intercept** | 1 | -2.3350 | 0.5111 | 20.8713 | <.0001 |
| **Age** | 1 | 0.0293 | 0.0132 | 4.9455 | 0.0262 |
| **Sector** | 1 | 1.6734 | 0.4873 | 11.7906 | 0.0006 |

Figure 1: Sample logistic regression output.

**Can you think of an example where we might want to make the probability threshold for predicting a "1" be less than 0.5?**

Any example where the consequence of a false positive (predicting one when the answer is in fact zero) is much less costly than a false negative (predicting 0 when the answer is in fact 1).

Example, predicting someone has COVID-19 given demographic variables. It may be useful to make threshold for a 1 prediction less than 50% because it is less costly to quarantine someone who is not actually infected than it is to not quarantine someone who is infected.

**Using non-technical terms, how might you describe an outlier in logistic regression?**

Observing a positive (or negative) result when the chance of observing a positive (or negative) result is extremely low.