

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?

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

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