Caijun Qin
11/02/2021

1. a.

$\log\left(\frac{\hat{\pi}_D}{\hat{\pi}_I}\right) = 3.3 - 0.2x$

$\log\left(\frac{\hat{\pi}_R}{\hat{\pi}_I}\right) = 1.0 + 0.3x$

$\log\left(\frac{\hat{\pi}_R}{\hat{\pi}_D}\right) = \log\left(\frac{\frac{\hat{\pi}_R}{\hat{\pi}_I}}{\frac{\hat{\pi}_D}{\hat{\pi}_I}}\right) = \log\left(\frac{\hat{\pi}_R}{\hat{\pi}_I}\right) - \log\left(\frac{\hat{\pi}_D}{\hat{\pi}_I}\right) = (1.0 + 0.3x) - (3.3 - 0.2x)$

$= -2.3 + 0.5x$

The slope, 0.5, can be interpreted the log of odds ratio increases (slope is positive) by 0.5 per unit increase ($10k) in income. The odds ratio here refers to the probability of being a Republican over a Democrat. Rather than log of odds ratio, an equivalent way to express the interpretation is that the odds of being Republican rather than the odds of being Democrat increases by $e^{0.5}$ per $10k increase in income.

b. $\hat{\pi}_R > \hat{\pi}_D \Rightarrow \frac{\hat{\pi}_R}{\hat{\pi}_D} > 1 \Rightarrow \log\left(\frac{\hat{\pi}_R}{\hat{\pi}_D}\right) > 0 \Rightarrow -2.3 + 0.5x > 0 \Rightarrow x > 4.6$

The domain of x that satisfies the condition can be expressed as $(4.6, \infty)$.

c.

Since Democrats, Republicans, and Independents together comprise all categories available for the response variable, the sum of their probabilities, even if estimated from a model, is 1.

$1 = \hat{\pi}_D + \hat{\pi}_R + \hat{\pi}_I = \frac{e^{3.3-0.2x}}{1+e^{3.3-0.2x}+e^{1.0+0.3x}} + \frac{e^{1.0+0.3x}}{1+e^{3.3-0.2x}+e^{1.0+0.3x}} + \hat{\pi}_I$

$\Rightarrow \hat{\pi}_I = 1 - \frac{e^{3.3-0.2x}+e^{1.0+0.3x}}{1+e^{3.3-0.2x}+e^{1.0+0.3x}} = \frac{1}{1+e^{3.3-0.2x}+e^{1.0+0.3x}}$

16.

Please reference the code attached. From experimenting in code, both models, one using the beginning LDL levels as a quantitative variable and the other treating that predictor as nominal only, seem to suggest that the treatment does have statistically significant influence on ending LDL levels past a fixed measuring threshold (from a frequentist approach). The L.R.T. from the VGAM package produced p-values of 4.074e-15 and 3.314e-16, respectively, with the null model only using the beginning LDL levels as the sole predictor whilst leaving out the type of cereal (control or treatment). With p-values smaller than the conventionally accepted alpha levels of 0.05 and even 0.01, we can reject the null hypothesis stating that whether psyllium was added or not i.e. the type of cereal had negligible effect on the ending LDL levels. From both fitted multinomial GLM models, log-likelihood values can be extracted to compute the AIC. In the respective order, these values are 322.5282 and 174.7082. Having both a smaller p-value from L.R.T. and a lower AIC value suggests that the second model, which sees beginning LDL levels as nominal only, is preferable over the first model. BIC posits another evaluation metric, but it is suitable to compare models with a large number of parameters and was thus omitted.