

6.

a. The study entails a general linear model with the following components.

- Random component, describing the distribution of residuals following a binomial distribution.

- Systematic component of $3.1870 - 0.5901X$.

- Link function of $g(\pi) = \ln \frac{\pi}{1-\pi} = 3.1870 - 0.5901X$, where $\mu = E[Y]$ at particular values taken by X . This can be interpreted as the proportion of people at the same level of being conservative who affiliate with the Democratic Party.

The link function can be re-arranged to an equation that predicts the proportion π given X .

$$\begin{aligned}\ln \frac{\pi}{1-\pi} &= 3.1870 - 0.5901X \\ \Rightarrow \frac{\pi}{1-\pi} &= e^{3.1870-0.5901X} \\ \Rightarrow \pi &= e^{3.1870-0.5901X} - e^{3.1870-0.5901X}\pi \\ \Rightarrow \pi(1 + e^{3.1870-0.5901X}) &= e^{3.1870-0.5901X} \\ \Rightarrow \pi(X) &= \frac{e^{3.1870-0.5901X}}{1+e^{3.1870-0.5901X}}\end{aligned}$$

With a negative estimated coefficient for the only predictor variable X , the predicted proportion of subjects being Democrats decreases towards zero (0) as X , the level of conservativeness, increases. Similarly, as X decreases, π increases towards one (1).

b. The confidence interval (Wald) with a level of 95% is (-0.896644, -0.283556). This is narrower and actually an enclosed subset of the profile likelihood interval computed previously (see code).

c.

The Wald test statistic is -3.772272, and the associated p-value is 0.0001617676. The extremely miniscule p-value indicates strong evidence to reject a null hypothesis stating that the true parameter β is 0. This can be interpreted more realistically as the ideological scale (how conservative someone is) does have an effect on the proportion of each group of people at the same level of conservative being Democrats. In fact, the negative value of the statistic suggests a negative relationship between the predictor and outcome ratio. Increasing the level of conservatism decreases the ratio of Democrats.

d.

For the Likelihood Ratio Test (LRT) produced a test statistic of 17.009 with a p-value of 3.72e-05. Again, the miniscule p-value provides strong evidence against stating that the full model and reduced model are statistically indifferent. The reduced model here just eliminates the predictor variable for ideological scale and leaves only the intercept behind in the model. One formula of computing the statistic is $2(L_1 - L_0)$, where capital L s refer to the maximum log-likelihoods for the full and reduced models, respectively. A large statistic of 17.009 (large in the sense of interpreting the tiny p-value) means that L_1 is much larger than L_2 , and therefore the single predictor should be considered relevant to the model.

Mini Project Findings

Please refer to computations in the attached R code. With an analysis of residual plots for both the Poisson and binomial random component, both plots seem to show directionality and unequal variance of residuals. A Poisson random component was decided for this case. Using the LRT and Wald test, a reduced model without the age midpoint predictor (just intercept) was compared to the full model with the null hypothesis that the true parameter of the age midpoint being zero (0). This was ultimately rejected when both tests rendered a tiny p-value of $2.2e-16$. Therefore, age range having an influential effect on nonmelanoma cancer rate cannot be simply ignored. Next, overdispersion was checked with both custom code and functions from packages, all of which confirm the existence of overdispersion strongly. To remedy this, the GLM model was redone with a quasipoisson (due to having larger variance than mean) and negative binomial random components, of which the latter performed better at reducing overdispersion by lowering the ration between chi-square statistic and the degrees of freedom for the model.