Test Exercise 5

Jorge Pineño

Theoretical Questions

We consider the logit model for direct mailing response:

$$\Pr[\text{resp}_i = 1] = \frac{\exp(\beta_0 + \beta_1 \cdot \text{male}_i + \beta_2 \cdot \text{active}_i + \beta_3 \cdot \text{age}_i + \beta_4(\text{age}_i/10)^2)}{1 + \exp(\beta_0 + \beta_1 \cdot \text{male}_i + \beta_2 \cdot \text{active}_i + \beta_3 \cdot \text{age}_i + \beta_4(\text{age}_i/10)^2)}$$

(a) Show that $\frac{\partial \Pr[\mathbf{resp}_i=1]}{\partial \mathbf{age}_i} + \frac{\partial \Pr[\mathbf{resp}_i=0]}{\partial \mathbf{age}_i} = 0$

Let $p_i = \Pr[\text{resp}_i = 1]$. Then $\Pr[\text{resp}_i = 0] = 1 - p_i$. Differentiating each with respect to age_i:

$$\frac{\partial \Pr[\text{resp}_i = 0]}{\partial \text{age}_i} = \frac{\partial(1 - p_i)}{\partial \text{age}_i} = -\frac{\partial p_i}{\partial \text{age}_i}$$

Hence,

$$\frac{\partial p_i}{\partial \text{age}_i} + \left(-\frac{\partial p_i}{\partial \text{age}_i}\right) = 0$$

(b) Show that recoding $\text{resp}_i^{\text{new}} = -\text{resp}_i + 1$ changes the sign of all parameters using the odds ratio

In the original model, the log-odds is:

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 \cdot \text{male}_i + \beta_2 \cdot \text{active}_i + \beta_3 \cdot \text{age}_i + \beta_4(\text{age}_i/10)^2$$

If we redefine the dependent variable to represent a negative response as 1 and a positive as 0, then:

$$\frac{\Pr[\text{resp}_i = 0]}{\Pr[\text{resp}_i = 1]} = \exp\left(-(\beta_0 + \beta_1 \cdot \text{male}_i + \beta_2 \cdot \text{active}_i + \beta_3 \cdot \text{age}_i + \beta_4(\text{age}_i/10)^2)\right)$$

Taking logs:

$$\log\left(\frac{1 - p_i}{p_i}\right) = -\left(\beta_0 + \beta_1 \cdot \text{male}_i + \beta_2 \cdot \text{active}_i + \beta_3 \cdot \text{age}_i + \beta_4(\text{age}_i/10)^2\right)$$

Thus, all parameter signs are reversed.

(c) Extending the model to allow age effect to differ by gender

The current model includes a quadratic effect of age:

$$\log\left(\frac{p_i}{1 - p_i}\right) = \ldots + \beta_3 \cdot \text{age}_i + \beta_4 \cdot \left(\frac{\text{age}_i}{10}\right)^2$$

This implies the odds ratio peaks at:

$$\text{age}^* = -\frac{\beta_3}{2\beta_4} \cdot 10^2 = 50$$

To allow this optimal age to differ by gender, we can include interaction terms:

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 \cdot \text{male}_i + \beta_2 \cdot \text{active}_i + \beta_3 \cdot \text{age}_i + \beta_4 \cdot \left(\frac{\text{age}_i}{10}\right)^2$$
$$+ \gamma_1 \cdot (\text{male}_i \cdot \text{age}_i) + \gamma_2 \cdot \left(\text{male}_i \cdot \left(\frac{\text{age}_i}{10}\right)^2\right)$$

This extended model allows the linear and quadratic age effects to differ between males and females.