Question 1

The counts of people died in 2009 in a certain state for different life length, genders, and races were (in 1000s):

<table>
<thead>
<tr>
<th>Life length</th>
<th>White males</th>
<th>Black males</th>
<th>White females</th>
<th>Black females</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-20</td>
<td>2.4</td>
<td>3.6</td>
<td>1.6</td>
<td>2.7</td>
</tr>
<tr>
<td>20-40</td>
<td>3.4</td>
<td>7.5</td>
<td>1.4</td>
<td>2.9</td>
</tr>
<tr>
<td>40-60</td>
<td>21.3</td>
<td>33.3</td>
<td>12.1</td>
<td>20.7</td>
</tr>
<tr>
<td>60+</td>
<td>72.9</td>
<td>55.6</td>
<td>84.9</td>
<td>73.7</td>
</tr>
</tbody>
</table>

Suppose that we are interested in the effect of gender on life length. To utilize the ordering among the life length categories, the proportional odds model was fitted. The SAS code and output are provided in the Appendix.

(a) [10 points] Write down the fitted model.

Answer:

\[
\begin{align*}
\text{logit}(\hat{\pi}_{1}) &= -3.76 + \beta_{5} \text{Black} + (\beta_{74}) \text{Female} + (\beta_{76}) \text{Black} \times \text{Female} \\
\text{logit}(\hat{\pi}_{2}) &= -2.80 + \beta_{5} \text{Black} + (\beta_{74}) \text{Female} + (\beta_{76}) \text{Black} \times \text{Female} \\
\text{logit}(\hat{\pi}_{3}) &= -2.99 + \beta_{5} \text{Black} + (\beta_{74}) \text{Female} + (\beta_{76}) \text{Black} \times \text{Female} \\
\text{Black} &= \begin{cases} 1 & \text{if Race} = \text{Black} \\ 0 & \text{otherwise} \end{cases} \quad \text{Female} = \begin{cases} 1 & \text{if Sex} = \text{Female} \\ 0 & \text{otherwise} \end{cases}
\end{align*}
\]

(b) [10 points] Based on the fitted model, what is the most likely life length group for a white male?

For white males,

\[
\begin{align*}
\text{logit}(\hat{\pi}_{1}) &= \hat{\beta}_{1} \\
\text{logit}(\hat{\pi}_{2}) &= \hat{\beta}_{2} \\
\text{logit}(\hat{\pi}_{3}) &= \hat{\beta}_{3} \\
\end{align*}
\]

So, \( \hat{\pi}_{1} = \frac{e^{\hat{\beta}_{1}}}{1 + e^{\hat{\beta}_{1}}} \approx 0.011 \) 

\( \hat{\pi}_{2} = \frac{e^{\hat{\beta}_{2}}}{1 + e^{\hat{\beta}_{2}}} \approx 0.029 \) 

\( \hat{\pi}_{3} = \frac{e^{\hat{\beta}_{3}}}{1 + e^{\hat{\beta}_{3}}} \approx 0.177 \) 

\( \hat{\pi}_{4} = 1 - \hat{\pi}_{3} = 0.823 \)

Largest, so most likely group is 60+.
(c) [10 points] According to the fitted model, what is the effect of sex on the life length? Can we conclude women are more likely to live longer than men? Explain your reason.

Since Sex and Race interact, effect of sex depends on Race:

\[ \text{Race} = \text{Black}: \logit(\pi_{ij}) = \hat{\alpha}_j + (0.7 \hat{\beta}_1 - 0.06) \text{Female} + 0.75, \quad j = 1, 2, 3 \]

\[ = (\hat{\alpha}_j + 0.75) - 0.80 \text{Female} \]

The odds of shorter life length (vs. longer) for female blacks are \( e^{-0.80} = 0.45 \) times the odds for male blacks.

\[ \text{Race} = \text{White}: \logit(\pi_{ij}) = \hat{\alpha}_j - 0.74 \text{ Female}, \quad j = 1, 2, 3. \]

The odds of shorter life length (vs. longer) for female whites are \( e^{-0.74} = 0.48 \) times the odds for male whites.

Since 0.45, 0.48 both are less than 1, we can say females are more likely to live longer.

Answer:

No. The Score test has \( p \)-value < 0.0001,

so we reject the proportional odds assumption.

(d) [5 points] Is the proportional odds assumption valid here? Judge by a statistical test.

(e) [10 points] If the assumption fails, what alternative model would you recommend? Identify the model name and write down the model.

Answer: Baseline-Categories Logit Model.

Compare each life length group to the last one [60+].

\[ \log(\frac{\pi_{ij}}{\pi_{4j}}) = \alpha + \beta_1 \text{(Black)} + \beta_2 \text{(Female)} + \gamma \text{(Black)(Female)} \]

\( j = 1, 2, 3. \)
Question 2

The following table refers to applicants to graduate school of a certain university for the fall 2010 session. Admission decisions are presented by gender of applicant, for the 2 largest graduate departments. Denote the 3 variables by A = whether admitted, G = gender, and D = department.

<table>
<thead>
<tr>
<th>Department</th>
<th>Whether admitted, male</th>
<th>Whether admitted, female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>120</td>
</tr>
</tbody>
</table>

The loglinear model (AG, AD, DG) was fit for the data and its SAS code and output are shown in the Appendix. Answer the following questions based on the fitted model and output.

(a) [16 points] Write down the fitted model. Report the deviance and df values, and comment on the goodness of fit.

Answer: \[ \log(\lambda_{ijk}) = 4.5606 - 1.3675 I_{E=A=Yes} + (-4.602) I_{E=D=1} + 1.23 I_{E=G=male} + 1.7215 I_{E=A=Yes} I_{E=D=1} + 0.8896 I_{E=A=Yes} I_{E=G=male} + (-0.8760) I_{E=D=1} I_{E=G=male} \]

\[ (2) \]

Deviance = 0.0234 and DF = 1, since \( \frac{\text{Deviance}}{\text{DF}} \ll 1 \),

(2) we can conclude the model is a good fit.

(b) [10 points] Odds ratios are often used to measure the association between two binary variables. What is the (estimated) odds ratio between A and G? What is 95% C.I. for the odds ratio? Explain the C.I. in the context of the situation.

For \( k = 1, 2 \):

Answer: \[ \log(\theta_{AG(k)}) = \hat{A}_{G_{k}} = 0.8896 \Rightarrow \hat{\theta}_{AG(k)} = e^{0.8896} = 2.43 \]

95% C.I. for \( \theta_{AG} \) is \((1.4829, 1.2963) \Rightarrow 95\% \text{ C.I. for } \theta_{AG(k)} \) is \((e^{1.4829}, e^{1.2963}) = (1.62, 3.66) \)

(3) For a given department, the odds of being admitted for male applicants is 1.62 to 3.66 times the odds for female applicants, with 95% confidence level.
(c) [8 points] What is the predicted number of female applicants admitted by Department 2 in the next fall?

Answer: 
\[ \hat{\mu}_{122} = e^{3.1931} = 24.36 \times \begin{bmatrix} 24 \end{bmatrix} \]

(d) [5 points] What is the estimated probability of being admitted by Department 2 for a female applicant?

Answer: 
\[ \hat{\pi}_{122} = \frac{\hat{\mu}_{122}}{n_{122}} = \frac{24.36}{2496} = 0.9 \text{ or } .0293 \]

(e) [16 points] While treating A as the response, write down corresponding logit model. Also provide the estimates of all terms in your logit model. Hint: The intercept is the estimated logit(\(\pi\)) at the baseline where \(\pi\) is the probability of being admitted.

Answer: 
\[ \text{loglinear} \quad (A 67, AD, DG) \]
\[ \text{logit when A is the response} \quad D + G \]

\[ \text{logit} (\hat{\pi}_{jk}) = \alpha + \beta_1 \times I_{D=13} + \beta_2 \times I_{G=\text{male}} \]

where \[ \hat{\pi}_{jk} = P[A = \text{yes} | D = \delta, G = k] \]

\[ \hat{\beta}_1 = \hat{AD}_{11} = .7215 \quad \text{and} \quad \hat{\beta}_2 = \hat{AG}_{11} = .8896 \]

\[ \hat{\pi} = \text{logit}(\hat{\pi}_{122}) = \log\left(\frac{.203}{1-.203}\right) = -1.3675 \quad \text{(the coef. of A)} \]

=> fitted model is \[ \text{logit} (\hat{\pi}_{jk}) = -1.3675 + .7215 I_{D=13} + .8896 I_{G=\text{male}} \]