DISCUSSION 6

Coupon Data

In a study of the effectiveness of coupons offering a price reduction on a given product, 1,000 homes were selected at random. A packet containing advertising material and a coupon for the product were mailed to each home. The coupons offered different price reduction (5, 10, 15, 20 and 30 dollars), and 200 homes were assigned at random to each of the price reduction categories. The data in the file coupon.txt is shown below.

<table>
<thead>
<tr>
<th>Price</th>
<th>redeemed</th>
<th>not redeemed</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>30</td>
<td>170</td>
</tr>
<tr>
<td>10</td>
<td>55</td>
<td>145</td>
</tr>
<tr>
<td>15</td>
<td>70</td>
<td>130</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>137</td>
<td>63</td>
</tr>
</tbody>
</table>

1. plot the data into R, create a data frame from the data, then add a column with the total sample size and another columns with the proportion with redeemed coupons. Plot these proportions versus price.

2. Fit a logistic regression model to this data and report the coefficients. Is price significant? How can we interpret the estimated coefficient value for price? Check residual plots.

\[ \exp(0.096834) = 1.102 \]

Hence, the odd of a coupon being redeemed is estimated to increase by 10.2 percent with each one dollar increase in the coupon value.

3. On a single plot, include
   - the observed proportion of redeemed coupon versus price.
   - the total number of samples from each price, next to each point.
   - a curve of the estimated probability of coupon redemption versus price.

4. According to the model, at what price would we expect a 30 percent of coupon redemption?

\[ \log \left( \frac{0.3}{0.7} \right) = -2.04435 + 0.096834 \times \text{price} \]

5. Use the model to predict the proportion of redeemed 25 dollar coupon.

\[ \hat{\pi} = \frac{\exp(-2.04435 + 0.096834 \times \text{price})}{1 + \exp(-2.04435 + 0.096834 \times \text{price})} \]

6. Based on the deviance, is the model fitting this data well?
   - Lower deviance means better fit to data.
   - Residual deviance is approximately \( \chi^2 \) when total \( n_i \) generally is big.
   - When an informative predictor is added to a model, we expect deviance to decrease a lot.
Solutions

# 1
coupon <- read.table('coupon.txt', header=T)
coupon$total = coupon$redeemed + coupon$notredeemed
coupon$prop = coupon$redeemed / coupon$total
plot(prop ~ price, coupon)

# 2
fit = glm(prop ~ price, weights=total, coupon, family=binomial)
# equivalently you can use
# fit = glm(cbind(redeemed, notredeemed) ~ price, coupon, family=binomial)

# 3
plot(prop ~ price, coupon)
within(coupon, text(x=price, y=prop, labels=as.character(total), pos=3))
price.grid = seq(3, 35, by =0.1)
newdata = data.frame(price = price.grid)
pred.grid = predict(fit, newdata, type="response")
lines(price.grid, pred.grid)

# 4
log(0.3/0.7) # -0.8472979
# therefore we solve -2.044348 + 0.09683363*price = -0.847 for price
-(-2.044348 - log(0.3/0.7)) / 0.09683363 # 12.36193

# 5
newprice = data.frame(price=25)
predict(fit, newprice, type="response")

# verify
(exp(-2.044348+0.09683363*25)/(1+exp(-2.044348+0.09683363*25)))

# 6
summary(fit)
pchisq(2.1668, df=3, lower.tail=F) # 0.5385173