3.1 We consider three nested regression models for modelling the number of car accidents according to region (`region`). The variables risk class (`risk`) has 3 categories and the number of years of driving experience (`exp`) is broken down into 4 categories.

Model	variables	$p+1$	$\ell(\widehat{\boldsymbol{\beta}})$	AIC	BIC
M_1	`risk`	3	-244.566	495.132	510.362
M_2	`risk + region`	$\star$	-151.620	$\star$	$\star$
M_3	`risk + region + exp`	10	-139.734	299.468	350.235

Table 1: Goodness-of-fit measures for three nested regression models with the number of parameters in each model $(p+1)$, the value of the log-likelihood function evaluated at the maximum likelihood estimate $(\ell(\widehat{\boldsymbol{\beta}}))$ and information criteria.

What is the difference between AIC and BIC of Model M_2 (in absolute value)?

Solution
The difference is approximately 35.54. The only information required is the number of parameters of M_2 and the sample size. We can use the first line to compute the latter, since

$$\text{BIC} = -2\ell(M_1) + 3\ln(n)$$

and we find upon solving the equation $n = 1184$. It only remains to compute the number of levels for the region variable. This is $\text{DF}(M_3) - \text{DF}(M_1) - 3 = 7$; there are $K = 4$ categories for years of experience, but only three parameters. The difference $\text{BIC} - \text{AIC} = 7 \cdot \{\ln(1184) - 2\}$.

3.2 A random variable X follows a geometric distribution with parameter p if its probability mass function is

$$\mathsf{P}(X = x) = (1-p)^{x-1}p, \qquad x = 1, 2, \ldots$$

(a) Write the likelihood and the log-likelihood of the n sample.
(b) Derive the maximum likelihood estimator for the parameter p.
(c) Compute the observed information matrix.
(d) Suppose we have a sample of 15 observations, $\{5, 6, 3, 7, 1, 2, 11, 8, 7, 34, 1, 7, 10, 1, 0\}$, whose sum is 103. Compute the maximum likelihood estimate and its approximate standard error.
(e) Compute the likelihood ratio and the Wald test statistics. Perform a test at level 5% for $\mathscr{H}_0 : p_0 = 0.1$ against the two-sided alternative $\mathscr{H}_a : p_0 \neq 0.1$.

Solution
(a)

$$L(p; \boldsymbol{x}) = \prod_{i=1}^{n}(1-p)^{x_i-1}p = (1-p)^{\sum_{i=1}^{n}(x_i-1)}p^n$$

$$\ell(p; \boldsymbol{x}) = \ln(1-p)\sum_{i=1}^{n}(x_i-1) + n\ln(p)$$

(b)

$$\frac{\mathrm{d}}{\mathrm{d}p}\ell(p; \boldsymbol{x}) = -\frac{1}{(1-p)}\sum_{i=1}^{n}(x_i-1) + \frac{n}{p}$$

Setting the score to zero and re-arranging the expression, we find

$$\frac{1}{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Since the second derivative is negative, the maximum likelihood estimator is the reciprocal mean, $\widehat{p} = \overline{X}^{-1}$.

(c) The observed information function is the negative of the hessian matrix. We have

$$\frac{\mathrm{d}^2}{\mathrm{d}p^2} \ell(p; \boldsymbol{x}) = -\frac{n(\overline{x} - 1)}{(1 - p)^2} - \frac{n}{p^2}$$

and the observed information evaluated at the maximum likelihood estimate is $j(\widehat{p}) = n\overline{x}^3 / (\overline{x} - 1)$.

(d) The maximum likelihood estimate is 0.1456 and its standard error is 0.0347.

(e) The Wald statistic and the likelihood ratio statistic equal

$$W = (\widehat{p} - p_0)^2 / \mathsf{Va}(\widehat{p}) = (0.1456 - 0.1)^2 / 0.0347 = 1.72$$
$$R = 2\{\ell(\widehat{p}) - \ell(p_0)\} = 2\{-45.11 + 45.39\} = 0.558.$$

Both can be compared to χ_1^2: the respective p-values are 0.19 and 0.45, so we fail to reject the null $\mathscr{H}_0 : p = 0.1$.

3.3 We consider a simple Poisson model for the number of daily sales in a store, which are assumed independent from one another. Your manager tells you the latter depends on whether the store is holding sales or not. The mass function of the Poisson distribution is

$$\mathsf{P}\left(Y_i = y_i \mid \texttt{sales}_i\right) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, \dots$$

and we model $\lambda_i = \exp(\beta_0 + \beta_1 \texttt{sales}_i)$, where $\texttt{sales}_i$ is a binary indicator equal to unity during sales and zero otherwise.

(a) Derive the maximum likelihood estimator of (β_0, β_1). *Hint: maximum likelihood estimators are invariant to reparametrization.*

Solution

We use the subscript s to denote events for which $\texttt{sales} = 1$ and r if $\texttt{sales} = 0$. We have $n = 12$, $n_r = 7$, $n_s = 5$. Let λ_s (λ_r) denote the expected number of daily sales during (outside of) sales and $\overline{y}_s$, $\overline{y}_r$ and $\overline{y}$ denote the average daily number of sales when $\texttt{sale} = 0$, $\texttt{sale} = 1$ and the overall sample mean. The easiest way to derive the maximum likelihood estimator is to obtain the maximum likelihood estimators $\widehat{\lambda}_r$ and $\widehat{\lambda}_s$, which are the average number of sales. By invariance of maximum likelihood estimator

$$\widehat{\lambda}_r = \exp(\widehat{\beta}_0), \qquad \widehat{\lambda}_s = \exp(\widehat{\beta}_0 + \widehat{\beta}_1).$$

We thus have $\widehat{\beta}_0 = \ln(\overline{y}_r)$ and $\widehat{\beta}_1 = \ln(\overline{y}_s / \overline{y}_r)$.

A more direct (and tedious) way to obtain $\widehat{\boldsymbol{\beta}}$ is by differentiating the log likelihood, setting the gradient to zero

and solving simultaneously for $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1\texttt{sales}_i) - \exp(\beta_0)\sum_{i=1}^{n}\exp(\beta_1\texttt{sales}_i) - \sum_{i=1}^{n}\ln(y_i!)$$

$$\frac{\partial\ell}{\partial\beta_0} = \sum_{i=1}^{n} y_i - \exp(\beta_0)\sum_{i=1}^{n}\exp(\beta_1\texttt{sales}_i) = n\overline{y} - \exp(\beta_0)\{n_r + n_s\exp(\beta_1)\}$$

$$\frac{\partial\ell}{\partial\beta_1} = \sum_{i=1}^{n} y_i\texttt{sales}_i - \sum_{i=1}^{n}\texttt{sales}_i\exp(\beta_0 + \beta_1\texttt{sales}_i) = n_s\overline{y}_s - \exp(\beta_0)n_s\exp(\beta_1)$$

$$\frac{\partial^2\ell}{\partial\beta_0^2} = -\sum_{i=1}^{n}\exp(\beta_0 + \beta_1\texttt{sales}_i)$$

$$\frac{\partial^2\ell}{\partial\beta_1^2} = -\sum_{i=1}^{n}\texttt{sales}_i\exp(\beta_0 + \beta_1\texttt{sales}_i) = -n_s\exp(\beta_0 + \beta_1)$$

$$\frac{\partial^2\ell}{\partial\beta_1\partial\beta_0} = -n_s\exp(\beta_0 + \beta_1)$$

We find $\exp(\beta_0) = \overline{y}_s/\exp(\beta_1)$, so substituting this term in the expression for $\partial\ell/\partial\beta_0 = 0$ and noting $n\overline{y} = n_r\overline{y}_r + n_s\overline{y}_s$, we get

$$n\overline{y}\exp(\beta_1) = \overline{y}_s\{n_r + n_s\exp(\beta_1)\}$$

and rearranging terms gives

$$\frac{n_r\overline{y}_s}{n\overline{y} - n_s\overline{y}_s} = \frac{\overline{y}_s}{\overline{y}_r} = \exp(\beta_1).$$

(b) Calculate the maximum likelihood estimates for a sample of size 12, where the number of transactions outside sales is $\{2; 5; 9; 3; 6; 7; 11\}$, and during sales, $\{12; 9; 10; 9; 7\}$.

Solution

Using software or the formulae, we get $\widehat{\beta}_0 = 1.8153$ and $\widehat{\beta}_1 = 0.4254$.

(c) Calculate the observed information matrix and use the latter to derive standard errors for $\widehat{\beta}_0$ and $\widehat{\beta}_1$ and a 95% confidence interval for the parameters.

Solution

The expression for the hessian simplifies when we substitute $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$: we get

$$j(\widehat{\boldsymbol{\beta}}) = \begin{pmatrix} n\overline{y} & n_s\overline{y}_s \\ n_s\overline{y}_s & n_s\overline{y}_s \end{pmatrix} = \begin{pmatrix} 90 & 47 \\ 47 & 47 \end{pmatrix}$$

Inverting this 2×2 matrix and taking the square root of the diagonal elements of $j^{-1}(\widehat{\boldsymbol{\beta}})$ gives the standard errors of $\widehat{\boldsymbol{\beta}}$, namely $\mathrm{se}(\widehat{\beta}_0) = 43^{-1/2} = 0.1525$ and $\mathrm{se}(\widehat{\beta}_1) = (90/43\cdot47)^{1/2} = 0.2110$. The 95% confidence intervals for β_0 and β_1 are respectively $[1.5164, 2.1142]$ and $[0.0118, 0.839]$.

(d) Your manager wants to know if the daily profits during sales period are different from those outside of the sales period. She calculates that the average profit during sales is \$20 per transaction, compared to \$25 normally. Test this hypothesis using a likelihood ratio test. *Hint: write the null hypothesis of equal profit in terms of the model parameters β_0 and β_1.* (**difficult**)

Solution

We want to test if the total of the average daily sales is the same during sales period, which amounts to testing for $\mathcal{H}_0 : 20\exp(\beta_0 + \beta_1) = 25\exp(\beta_0)$, or $\beta_1 = \log(5/4)$. We can compute the restricted maximum log likelihood,

viz.

$$\ell_{\mathrm{p}}(\beta_0) = \sum_{i=1}^{n} y_i \{\beta_0 + \log(5/4)\mathtt{sales}_i\} - \exp(\beta_0)\sum_{i=1}^{n}(5/4)^{\mathtt{sale}_i}$$

Computing the derivative and setting it to zero gives $n\overline{y} - \exp(\beta_0)(n_r + 5n_s/4) = 0$ and so $\widehat{\beta}_0 = \ln(n\overline{y}) - \ln(n_r + 5n_s/4) = 1.9158$. We can compute the maximum log likelihood under both null and alternative hypothesis. We get $\ell(\widehat{\boldsymbol{\beta}}) = 93.37082$ and $\ell_{\mathrm{p}}(\widehat{\beta}_0) = 92.91$, so the likelihood ratio statistic is $R = 0.92$ and we compare this to a χ_1^2 null distribution, whose 0.95 quantile is 3.84; we fail to reject the two-sided null hypothesis that profits are the same regardless of whether there is a sale occuring. The p-value is 0.337.