MA50260 Statistical Modelling

Lecture 11: GLM - Hypothesis Test and Model Comparison

Ilaria Bussoli

March 12, 2024

Motivation

We have explored how the `glm` function in `R` estimates the regression coefficients $\beta_1, \ldots, \beta_p$.

This led us to iteratively re-weighted least squares (IRWLS).

Today we consider

- Testing for significance
- Performing model selection

Distribution of Maximum Likelihood Estimator

The maximum likelihood estimator is asymptotically normal,

$$\underline{\hat{\theta}}(\mathbf{Y}) \sim \mathrm{MVN}\left(\underline{\theta}\,, \mathcal{I}^{-1}(\underline{\theta})\right).$$

Distribution of Maximum Likelihood Estimator

The maximum likelihood estimator is asymptotically normal,

$$\underline{\hat{\theta}}(\mathbf{Y}) \sim \mathrm{MVN}\left(\underline{\theta}, \mathcal{I}^{-1}(\underline{\theta})\right).$$

For the MLE of a GLM, we have

$$\underline{\hat{\beta}}(\mathbf{Y}) \sim \mathrm{MVN}_p\left(\underline{\beta}, \phi(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\right),$$

where

$$W_{ii} = \frac{1}{V(\mu_i)g'(\mu_i)^2} \quad (i = 1, \ldots, n).$$

We use this to constructs confidence interval and for hypothesis testing.

Note, we use $\hat{\mu}_i$ to estimate W_{ii}.

Case 1: ϕ is known

Consider

$$H_0 : \beta_j = b \qquad \text{vs.} \qquad H_1 : \beta_j \neq b.$$

The **test statistic** is

$$z = \frac{\hat{\beta}_j - b}{\sigma_j},$$

where

$$\sigma_j = \sqrt{\phi(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}_{j,j}}.$$

We compare $|z|$ to the $(1 - \alpha/2) \times 100\%$ quantile of the standard normal distribution.

The $(1 - \alpha) \times 100\%$ **confidence interval** is calculated as

$$\hat{\beta}_j \pm z_{1-\alpha/2}\, \sigma_j.$$

Case 2: ϕ is unknown

We estimate

$$\hat{\sigma}_j = \sqrt{\hat{\phi}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}_{j,j}}.$$

The **test statistic** is

$$t = \frac{\hat{\beta}_j - b}{\hat{\sigma}_j},$$

and we compare $|t|$ to the $(1 - \alpha/2) \times 100\%$ quantile of the t_{n-p}-distribution.

The $(1 - \alpha) \times 100\%$ confidence interval is

$$\hat{\beta}_j \pm t_{n-p}(1 - \alpha/2)\, \hat{\sigma}_j.$$

Example - Beetle Mortality

```r
beetle_glm <- glm(cbind(dead,alive) ~ dose,
                  family = binomial, data = beetles)
summary(beetle_glm)$coefficients
```

```
##              Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -2.058899 0.44315258 -4.646027 3.383879e-06
## dose         0.207153 0.03007281  6.888382 5.643066e-12
```

- The p-value for β_2 is close to zero $\Rightarrow$ Dose is significant.
- A 95% confidence interval for β_2 is

$$\hat{\beta}_2 \pm z_{1-\alpha/2}\sqrt{\sigma_2^2} = 0.207 \pm 1.96 \times 0.03 = (0.148, 0.266).$$

Saturated Models

When can we say that a GLM fits perfectly?

Saturated Models

When can we say that a GLM fits perfectly?

Recall that $g[b'(\theta_i)] = \mathbf{x}_i^{\mathrm{T}}\underline{\beta}$ and consider the log-likelihood for observation i,

$$\ell_i = \ell(\theta_i, \phi \mid y_i) = w_i \frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi),$$

which is maximised for $b'(\theta_i) = y_i$.

Saturated Models

When can we say that a GLM fits perfectly?

Recall that $g[b'(\theta_i)] = \mathbf{x}_i^{\mathrm{T}} \underline{\beta}$ and consider the log-likelihood for observation i,

$$\ell_i = \ell(\theta_i, \phi \mid y_i) = w_i \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi),$$

which is maximised for $b'(\theta_i) = y_i$.

We call a model **saturated** if $b'(\theta_i) = y_i$ for all $i = 1, \ldots, n$.

In general, we need one parameter for each observation for a saturated model.

The parameters are $\tilde{\theta}_1, \ldots, \tilde{\theta}_n$, with $\tilde{\theta}_i = b'^{-1}(y_i)$.

Deviance

To assess model fit, we consider the **deviance**:

$$D = 2\left[\ell\left(\underline{\tilde{\theta}}\right) - \ell\left(\underline{\hat{\theta}}\right)\right]\phi,$$

where $\underline{\hat{\theta}}$ is the MLE for $\underline{\theta}$.

If we use a distribution from the **exponential family**, then

$$D = \sum_{i=1}^{n} 2w_i\left[y_i\left(\tilde{\theta}_i - \hat{\theta}_i\right) - b\left(\tilde{\theta}_i\right) + b\left(\hat{\theta}_i\right)\right].$$

The **scaled deviance** is defined as

$$D^* = D/\phi = 2\left[\ell\left(\underline{\tilde{\theta}}\right) - \ell\left(\underline{\hat{\theta}}\right)\right].$$

Model Comparison (I)

Consider the models $\mathcal{M}_1$ and $\mathcal{M}_2$, with $\mathcal{M}_1$ being nested in $\mathcal{M}_2$.

We want to test

$$\begin{aligned} H_0 &: \quad \mathcal{M}_1 \text{ adequately describes } \mathbf{y} \\ H_1 &: \quad \mathcal{M}_2 \text{ is required.} \end{aligned}$$

Model Comparison (I)

Consider the models $\mathcal{M}_1$ and $\mathcal{M}_2$, with $\mathcal{M}_1$ being nested in $\mathcal{M}_2$.

We want to test

$$\begin{aligned} H_0 : &\quad \mathcal{M}_1 \text{ adequately describes } \mathbf{y} \\ H_1 : &\quad \mathcal{M}_2 \text{ is required.} \end{aligned}$$

The deviance (likelihood ratio test) statistic is

$$D(\mathcal{M}_2, \mathcal{M}_1) = 2\left[\ell\left(\underline{\hat{\beta}}^{(2)}\right) - \ell\left(\underline{\hat{\beta}}^{(1)}\right)\right].$$

We refer to $\mathcal{M}_2$ and $\mathcal{M}_1$ as the **full** and **reduced** model, respectively.

Model Comparison (II) - ϕ is known

Under H_0, asymptotically as $n \to \infty$,

$$D(\mathcal{M}_2, \mathcal{M}_1) \sim \chi^2_{p_2 - p_1}.$$

We can also relate it to the scaled deviances since

$$\begin{aligned} D(\mathcal{M}_2, \mathcal{M}_1) &= 2\left[\ell\left(\underline{\hat{\beta}}^{(2)}\right) - \ell\left(\underline{\hat{\beta}}^{(1)}\right)\right] \\ &= 2\left[\ell\left(\underline{\hat{\beta}}^{(2)}\right) - \ell\left(\underline{\tilde{\beta}}\right) + \ell\left(\underline{\tilde{\beta}}\right) - \ell\left(\underline{\hat{\beta}}^{(1)}\right)\right] \\ &= D_1^* - D_2^* = \frac{D_1 - D_2}{\phi}. \end{aligned}$$

If ϕ is known, evaluate $D(\mathcal{M}_2, \mathcal{M}_1)$ and compare to the corresponding critical value, z_c^2 from the $\chi^2_{p_2 - p_1}$ distribution.

Model Comparison (III) - ϕ is unknown

Since $D_2^* \sim \chi^2_{n-p_2}$ asymptotically, we can estimate

$$\phi \approx D_2/(n - p_2).$$

Model Comparison (III) - ϕ is unknown

Since $D_2^* \sim \chi^2_{n-p_2}$ asymptotically, we can estimate

$$\phi \approx D_2/(n-p_2).$$

Consider the ratio

$$\frac{(D_1^* - D_2^*)/(p_2 - p_1)}{D_2^*/(n-p_2)} = \frac{(D_1 - D_2)/(p_2 - p_1)}{D_2/(n-p_2)} \sim F_{p_2-p_1, n-p_2}$$

and using that the ratio of two χ^2 distributions is an F distribution, specified by two degrees of freedom.

Note that this is still an approximation for all GLMs apart from the Gaussian model.