Binomial Regression II

Reminder: questions from Challenger disaster

- Forecast probability of an O-ring being damaged when the launch temperature is 29 ^oF.
- How good is our forecast? Can we provide a confidence interval?
- Is temperature useful to predict the O-ring failing?

Is temperature useful to predict the O-ring failing?

Y_i, the number of damaged O-rings on the i-th launch, has distribution

$$Y_i \sim \text{bin}(6, p_i)$$

where

$$\log p_i/(1 - p_i) = \eta_i = \beta_0 + \beta_1 t_i.$$

Test for association between the number of damaged O-rings and temperature:

$$H_0 : \beta_1 = 0 \quad H_a : \beta_1 \neq 0$$

- Wald Test
- Likelihood Ratio Test

Wald Test

Reminder: asymptotic normality of MLE

$$\hat{\theta}_i \sim asy.\ N(\theta_i^*, [\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1}]_{i,i}).$$

Test for association between the number of damaged O-rings and temperature

$$H_0 : \beta_1 = 0 \quad H_a : \beta_1 \neq 0$$

Wald test statistic:

$$z^* = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \sim asy.\ N(0,1) \text{ under } H_0.$$

Challenger disaster

- See R script and result in "Wald Test" of Challenger.pdf
- $|z^*| = 4.07 > 1.96$ (critical value $N(0,1)$ at $\alpha = 0.05$) $\Rightarrow$ reject H_0.
- p-value $= 0.0000476$.

Likelihood Ratio Test (LRT)

Test for association between the number of damaged O-rings and temperature

$$H_0 : \beta_1 = 0 \quad H_a : \beta_1 \neq 0$$

Full model (F):

- $\eta_i = \beta_0 + \beta_1 t_i$
- Maximum log likelihood: $\log \mathcal{L}(\hat{\boldsymbol{\beta}}^F)$
- $\hat{\boldsymbol{\beta}}^F$: MLE of the parameters in the full model.

Reduced model (R):

- $\eta_i = \beta_0$
- Maximum log likelihood: $\log \mathcal{L}(\hat{\boldsymbol{\beta}}^R)$
- $\hat{\boldsymbol{\beta}}^R$: MLE of the parameters in the reduced model.

Compare two models.

- Likelihood ratio test statistic:

$$LR^* = -2 \left[\log \mathcal{L}(\hat{\boldsymbol{\beta}}^R) - \log \mathcal{L}(\hat{\boldsymbol{\beta}}^F) \right] \sim asy.\ \chi^2_1 \text{ under } H_0.$$

- $LR^* >$ critical value from χ^2_1 at $\alpha \Rightarrow$ reject H_0.

Wald test vs Likelihood Ratio Test (LRT)

Wald test and LRT are asymptotically equivalent.

$$[z^*]^2 = \left[\frac{\hat{\beta}_1}{se(\hat{\beta}_1)}\right]^2 \sim asy.\ \chi^2_1$$

- Precisely speaking, two tests are asymptotically equivalent in the sense that under H_0 they reach the same decision with probability approaching 1 as n goes to infinite.

However, the chi-squared approximation to the log likelihood ratio is generally better than the normal approximation to the MLE.

Likelihood Ratio Test (LRT): Challenger disaster

- See R script and result in “Likelihood Ratio test” and “Wald Test vs Likelihood Ratio test” of Challenger.pdf
- $LR^* = 21.98 > 3.84$ (critical value from χ^2_1 at $\alpha = 0.05$) $\Rightarrow$ reject H_0.
- p-value $=$ 0.0000027.

Likelihood Ratio Test (LRT) for model selection

In general, likelihood ratio test is used to select between two nested models (one model can be obtained by constraining parameters of another model).

Full model (F):

- Maximum log likelihood: $\log \mathcal{L}(\hat{\boldsymbol{\theta}}^F)$

Reduced model (R):

- Maximum log likelihood: $\log \mathcal{L}(\hat{\boldsymbol{\theta}}^R)$

Let k indicate the difference in the number of parameters between two models.

Compare two nested models.

- Under the reduced model

$$LR^* = -2\left[\log \mathcal{L}(\hat{\boldsymbol{\theta}}^R) - \log \mathcal{L}(\hat{\boldsymbol{\theta}}^F)\right] \sim asy.\ \chi^2_k.$$

- $LR^* >$ critical value from χ^2_k at $\alpha \Rightarrow$ select the full model.

(Scaled) Deviance

The scaled deviance is used to judge model adequacy.

For the binomial regression model the deviance is the same as the *scaled deviance*, which is defined as the log likelihood ratio for the fitted model compared to the saturated model.

Full model (F):

- The saturated model has the same number of parameters and the observations.
- Maximum log likelihood: $\log \mathcal{L}(\hat{\boldsymbol{\theta}}^F)$

Reduced model (R):

- The fitted model.
- Maximum log likelihood: $\log \mathcal{L}(\hat{\boldsymbol{\theta}}^R)$

The scaled deviance:

$$D = -2\left[\log \mathcal{L}(\hat{\boldsymbol{\theta}}^R) - \log \mathcal{L}(\hat{\boldsymbol{\theta}}^F)\right].$$

(Scaled) Deviance

Warning: the number of parameters in the saturated model is n, which is not fixed, so the theory of maximum likelihood does not apply, and D may not converge to a chi-squared distribution.

(Scaled) Deviance for binomial regression model

The saturated model allows for one parameter for each observation. For binomial regression the saturated model has $p_1, p_2, \ldots, p_n$ as parameters. Clearly, for this model we estimate p_i by y_i/m_i. Let $\hat{p}_i = g^{-1}(\mathrm{x}_i^T \hat{\beta})$ be our (not saturated) model estimate of p_i, then the scaled deviance is

$$
\begin{aligned}
D &= -2 \sum_{i=1}^{n} \Big(y_i (\log \hat{p}_i - \log \frac{y_i}{m_i}) \\
&\quad + (m_i - y_i)(\log(1 - \hat{p}_i) - \log(1 - \frac{y_i}{m_i})) \Big) \\
&= -2 \sum_{i=1}^{n} \left(y_i \log \frac{\hat{y}_i}{y_i} + (m_i - y_i) \log \frac{m_i - \hat{y}_i}{m_i - y_i} \right)
\end{aligned}
$$

where $\hat{y}_i = m_i \hat{p}_i$ is the i-th fitted value.

(Scaled) Deviance for binomial regression model for testing model adequacy

It just happens: if $m_i p_i$ and $m_i(1 - p_i)$ are large enough (≥ 5 is a common rule of thumb), then for a binomial regression model, if the model is adequate then $D \approx \chi^2_{n-k}$, where k is the number of parameters in the fitted model (including β_0).

In this case the (scaled) deviance can be used as a test for model adequacy. If D is too large (as compared to a χ^2_{n-k}), then the model is missing something.

For a binomial model with small m_i we can't use the (scaled) deviance directly to test model adequacy, but we can still use it for model selection.

Use the scaled deviance for model selection (LRT)

If model A is nested within model B, and model A has (scaled) deviance D^A and model B has (scaled) deviance D^B, then

$$D^A - D^B = -2\left[\log \mathcal{L}(\hat{\boldsymbol{\theta}}^A) - \log \mathcal{L}(\hat{\boldsymbol{\theta}}^B)\right],$$

where $\hat{\boldsymbol{\theta}}^A$ and $\hat{\boldsymbol{\theta}}^B$ are MLEs for the models A and B, respectively.

That is, the log likelihood for the saturated model cancels, and we are left with the log likelihood ratio.

Use the scaled deviance for model selection (AIC)

The Akaike Information Criterion is used for model selection:

$$\text{AIC} = 2k - 2\log \mathcal{L}(\hat{\boldsymbol{\theta}})$$

where k is the number of parameters in the model. Given a choice, we prefer that model with the smaller AIC.

If model B has s more parameters than model A (not necessarily nested within B), then

$$\begin{aligned} \text{AIC}^B - \text{AIC}^A &= 2s - 2\log \mathcal{L}(\hat{\boldsymbol{\theta}}^B) + 2\log \mathcal{L}(\hat{\boldsymbol{\theta}}^A) \\ &= 2s - D^A + D^B. \end{aligned}$$

(Scaled) deviance: Challenger disaster

- See R script and result in “Deviance” of Challenger.pdf

Reminder: questions from Challenger disaster

- Forecast probability of an O-ring being damaged when the launch temperature is $29\,^oF$.
- How good is our forecast? Can we provide a confidence interval?
- Is temperature useful to predict the O-ring failing?

Learning goals

Understand binomial regression

- know when you should use binormal regression.
- be able to write binomial regression model and its likelihood.
- be able to obtain estimators of parameters or function of parameters using R script.
- be able to quantify uncertainty of the estimators (e.g., computing CI).
- be able to test hypothesis.
- be able to do model selection.

Understand asymptotic properties of MLEs (maximum likelihood estimators)

- use asymptotic normality of MLEs to quantify uncertainty of the estimators (e.g., Wald CI).

Undertand Wald test and likelihood ratio test (LRT)

- use them to test hypothesis in binomial regression

Understand (scaled) deviance

- use it to test model adequacy or perform LRT and model comparison.