STATS 111/202

Lecture 11: More GLM Inference and Model Comparisons

2/24/2025

Ana Maria Kenney

UC Irvine Department of Statistics

GLM Inference

4.2.3

Confidence interval review

- We are further interested in linear combination of the coefficients.

- Example: Confidence interval for $\beta_1 + 15\beta_3$ or $\beta_0 + 20\beta_1$.

- The linContr.glm() function will create 95% confidence intervals for all the odds ratios in the model that use a combination of the coefficients.

- Function has 3 inputs.
  - contr.names = The names of the coefficients to be used in the contrast.
  - contr.coef = The coefficients to use in forming the linear combination (the covariate values).
  - model = The name of the model to pull the estimates from.

GLM Inference

- Consider our current model:

$$\log\left(\mu / 1 - \mu\right) = \beta_0 + \beta_1 sbp + \beta_2 age + \beta_3 sbp \times age$$

- Let's say we want to interpret the effect of sbp for individuals at 20 years of age

- The effect of a 1 unit increase in sbp when age = 20, results in an estimated odds ratio of $e^{\widehat{\beta}_1 + \widehat{\beta}_3 20}$

- This **only provides a point estimate of the odds ratio**, we want to compute a confidence interval for $\hat{\beta}_1 + \hat{\beta}_3 20$ and exponentiate.

GLM Inference

- Recall, the general form of a CI:

$$\hat{\theta} \pm z_{\alpha/2} \, \widehat{SD}(\hat{\theta})$$

- Now we need to compute the $\mathrm{var}(\hat{\beta}_1 + \hat{\beta}_3 20)$

- Recall, if a and b are constants then we have:

$$var(aX + bZ) = a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\sigma_{12}$$

- Where $\sigma_1^2 = var(X), \sigma_2^2 = var(Z), and \; \sigma_{12} = cov(X, Z)$

- Thus, we need to compute these three components

GLM Inference

Back to our example:

- We can now compute the CI for $\beta_1 + \beta_3 20$ and need to calculate:
  $$var(\beta_1 + \beta_3 20) = var(\beta_1) + 20^2 var(\beta_3) + 2 \times 20 cov(\beta_1, \beta_3)$$

- We can estimate this variance-covariance matrix in R using the `vcov(model)` command

- Combine all 3 required elements in the sum and take the square root for the confidence interval.

CI contrasts example

Model comparisons

The saturated model:

- We define the saturated model to be the one that fits the observed data perfectly

- This requires that every observation in the dataset have its own specific coefficient

- There will be as many estimated parameters as there are observation points

- Therefore, the model will just end up estimating each response perfectly, but at the expense of computing too many estimates of coefficients, effectively using up all the degrees of freedom

Model comparisons

- Let L_S be the maximized value of the likelihood under the saturated model and let L_M be the maximized value of the likelihood under our current proposed model
  - We know that $L_M \leq L_S$

- Then we define deviance as:

$$Deviance = -2 \log\left(L_M/L_S\right) = -2[\log(L_M) - \log(L_S)]$$ where log is the natural log

- It compares the current model to the saturated one

Model comparisons

- Let L_S be the maximized value of the likelihood under the saturated model and let L_M be the maximized value of the likelihood under our current proposed model
  - We know that $L_M \leq L_S$

- Then we define deviance as:

$Deviance = -2\log\left(L_M/L_S\right) = -2[\log(L_M) - \log(L_S)]$ where log is the natural log

- It compares the current model to the saturated one

The lower the deviance, the closer the two models are to each other (lower is better for the current proposed model)

Model comparisons

Null model

- The null model is the model which has only a single parameter (intercept) and has a maximized value of the likelihood L_N
  - We fit this under the null hypothesis were all slope coefficients are equal to 0

- Then the null Deviance is:

$$Null\ Deviance = -2\log\left(L_N\big/L_S\right) = -2[\log(L_N) - \log(L_S)]$$

- It compares the null model to the saturated one

Model comparisons

- In R output, the null deviance is $-2 \log\left(L_N / L_S\right)$ and residual deviance is $-2 \log\left(L_M / L_S\right)$

- Null deviance is approximately chi-squared with degrees of freedom $df_n - df_s$ (df_s is degrees of freedom of saturated model and df_n is degrees of freedom of null model).

- Residual deviance is approximately chi-squared with degrees of freedom $df_m - df_s$ (df_s is degrees of freedom of saturated model and df_m is degrees of freedom of the current proposed model).

- Let n be the sample size of our dataset.
  - The degrees of freedom for the saturated model is 0
  - The degrees of freedom for the null model is n – 1, where n is the number of observations
  - The degrees of freedom for the proposed current model is n – (p + 1) = n – p – 1, where p is the number of slope coefficients

Model comparisons

- Can test the proposed model against the null model by looking at Null−Residual deviance (that is Null deviance minus Residual deviance) which has an approximate chi-squared distribution with degrees of freedom $df_n - df_n = p$ (where p is the number of slope coefficients in the proposed model).

- The idea is to see how far from the saturated model is the null model and the proposed model.

If null model and proposed model are approximately the same deviance from the saturated model, then we would have no evidence that the proposed model is better than the null model.

Model comparisons

- The formal test has null hypothesis H_0, that the null model holds (only intercept, all slope β's are 0). This means the null model fits the data well enough (compared to proposed model).

- The alternative is H_a, that the null model does not hold compared to proposed model. This means the null model does not fit the data well enough (compared to proposed model).

- The alternative will conclude the proposed model is better than the null model (while accounting for the difference in the degrees of freedom).

Model comparisons

- Our test statistic follows: Null Deviance – Residual Deviance

- This is approximately chi-squared distributed with degrees of freedom p (the number of slope coefficients in the proposed model).

- Small p-value will mean we reject the null and conclude evidence for the proposed model.

- If the p-value is large, we will tend to fail to reject the null. Concluding no evidence for the alternative (the null model is good enough).

Example

Model comparisons

- Now to test several coefficient equalling 0.

- Let the null follow $H_0: \beta_j = 0$ for some values of j vs H_a at least one is not 0

- For example $H_0: \beta_1 = \beta_3 = 0$ vs $H_0: \beta_1 \neq 0 \; or \; \beta_3 \neq 0$

- Will have a full model and a reduced model. Full model has no constraint (all original β's in it) and reduced model is fit under the null (some β's set to 0).

- Reduced model has k many coefficients and full model has p > k many coefficients.

Model comparisons

- Let L_1 be the maximized value of the likelihood under the full model and L_0 let be the maximized value of the likelihood under the reduced model

- Consider the test statistic:

$$D = -2 \log \left(L_0 / L_1 \right) = -2[\log(L_0) - \log(L_1)]$$

- Note that

$$D = -2[\log(L_0) - \log(L_S)] - (-2[\log(L_0) - \log(L_S)])$$

Where L_S was the maximized value of the likelihood under the saturated model.

- D has an approximate chi-squared distribution with degrees of freedom equal to $p - k$ (difference in the number of coefficients between full and reduced models).

- $D \sim \chi^2_{p-k}$, can use this to obtain p-value for test of H_0 against H_a

Model comparisons

Full vs reduced model:

- The test statistic is the reduced model deviance minus the full model deviance (just called deviance in our function in the next few slides).

- This test statistic follows an approximate chi squared distribution (the null hypothesis is that the reduced model is good enough and alternative is that the null is not true, which is to imply the full model is needed).

- The degrees of freedom (df) is the difference in the number of coefficients between the models. The number of coefficients in the full model minus the number of coefficients in the reduced model.

- P-value will be the area above our test statistic using a chi squared distribution with df degrees of freedom (in R, this is 1-pchisq(x, df)).

Likelihood ratio test example

General goodness of fit

Goodness of fit test for binary (logistic) regression

- Now to present a general test that was similar to the F-test that tested if all coefficients in a linear regression model are equal to 0 or not.

- In a logistic regression setting, this test is called the Hosmer and Lemeshow goodness of fit test.

- First we fit the model we have and generate fitted probabilities of event (or 1) for each sampling unit.

- Next, order the probabilities and group the data by quantiles of the fitted probabilities $\hat{\mu}$ (the number of groups, g, is specified by the researcher, the default is usually g=10).

General goodness of fit

- For each group, j=1,2,…,g, compute the observed and expected number of events (Y=1), o_j, e_j based on $\hat{\mu}$
- Note the total number of observations (Y=1 and Y=0) in each group is n_j and e_j (the expected count of 1's) is the sum of the probabilities in group j.
- Compared the observed counts to the expected counts across all g many groups by creating the appropriate chi-squared statistic.

$$\chi_{HL}^2 = \sum_{j=1}^{g} \frac{\left(o_j - e_j\right)^2}{e_j\left(1 - \frac{e_j}{n_j}\right)}$$

- If the fitted model is appropriate, the χ_{HL}^2 will have an approximate chi-square distribution with g-2 degrees of freedom.
- The test assesses whether or not the observed event rates match expected event rates in subgroups (based on their probabilities) of the model population.

General goodness of fit

- Let us conduct a goodness of fit test on the data we have been using so far.

- The Hosmer and Lemeshow test has null hypothesis, H_0 , that the model is fitting the observed data fairly well.

- This happens when the expected counts and the observed counts are close to one another.

- As a result, the chi-squared statistic will be small, and so p-value will be large.

- Again, the null is that the model is fitting the data well and the alternative is that the null is not true (model is not fitting the data well).

Goodness of fit test example