

Other Regression Analyses

Lecture 9-1

1 Outline

- Generalized Linear Models
- Logistic Regression
- Poisson Regression

2 Generalized Linear Models

2.1 Generalized Linear Models

The theory of linear models can be expanded to *generalized linear models* (GLM). The form of the generalized linear model involves introducing a link function $\eta(Y_j)$ such that

$$\eta(Y_j) = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + \varepsilon_j$$

! Important

The β terms are coefficients, with β_0 known as the intercept of the model. X_{1j} are predictor variables or covariates, also known as confounding variables. The term ε_j corresponds to the noise term of the model. This term does not contribute to the mean, rather this partition captures the variability in the data.

! Important

The function $\eta(Y_j)$ is referred to as a link function. A link function is slightly different from a transformation: Link functions are applied to the expected value, while a transformation function is applied to each data point.

2.2 Link Function Examples

Some examples of link functions include:

- Logit function $\eta(p_j) = \log(\frac{p_j}{1-p_j})$
- Probit function $\eta(p_j) = \Phi(p_j)$, $\Phi(p_j)$ being the cumulative distribution function of the standard normal distribution.
- Logarithmic function $\eta(Y_j) = \log Y_j$

3 Logistic Regression

3.1 Logistic Regression

Logistic regression methods are used when the response variable is dichotomous (Yes/No, 0/1).

Note

Logistic regression can be used for prediction and explanation of trends in the data. In explanatory analysis for dichotomous variables, the parameter of interest is the *probability of success*, p .

Logistic regression uses a logit link function such that:

$$\eta(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

3.2 Effect Sizes

The coefficient β_k for the k^{th} predictor in the linear model can be interpreted as the *natural logarithm of the odds ratios*.

Odds Ratios

The odds of an event is the ratio of the probability of success and the probability of failure. Specifically, the odds of an event with a success probability p is:

$$odds = p/(1-p)$$

The odds ratio compares the odds of an event for different groups or values of the predictor.

$$OR = odds_1 / odds_2 = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)} = e^{\beta_k}$$

3.3 R implementation

The function `glm()` in R can be used to perform logistic regression.

i Note

`glm()` can be used for different distributions (Poisson, Negative Binomial, etc.) The argument `family="binomial"` needs to be added in `glm()` to specify logistic regression.

3.3.1 Estimated Probabilities and Odds Ratios

The `emmeans()` function from the `emmeans` package can provide estimated probabilities for different groups.

Odds ratios can be calculated using the `contrast()` function in `emmeans()`.

3.3.2 Sample Code

```
mod1 <- glm(y~x, data=df, family="binomial")
summary(mod1) # outputs coefficients

library(emmeans)
em_mod_glm <- emmeans(mod1, ~x, type="response")
summary(em_mod_glm, infer=T)

con_mod_glm <- contrast(em_mod_glm, method="pairwise", type="response")
summary(con_mod_glm, infer=T)
```

4 Poisson Regression

4.1 Poisson Regression

Poisson regression is used primarily for modeling count data. Examples include bacterial counts, animal counts, number of patients who recovered/died.

i Note

Recall: Poisson distribution. In explanatory analysis for dichotomous variables, the parameter of interest is the *expected number of events for a unit of time and space*, λ . Poisson regression uses a logarithmic link function such that:

$$\eta(\lambda) = \log(\lambda) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

4.2 Effect Size

The coefficient β_k for the k^{th} predictor in the linear model can be interpreted as the *natural logarithm of the incidence rate ratios*.

i Note

λ is also known as the incidence rate in public health contexts. The odds ratio compares the incidence rate for different groups or values of the predictor.

$$IRR = \lambda_1 / \lambda_2 = e^{\beta_k}$$

4.3 R implementation

The function `glm()` in R can be used to perform Poisson regression.

i Note

The argument `family="Poisson"` needs to be added in `glm()` to specify Poisson regression.

4.3.1 Incidence Rates and Incidence Rate Ratios

The `emmeans()` function from the `emmeans` package can provide estimates for λ .

Incidence rate ratios can be calculated using the `contrast()` function in `emmeans()`.

4.3.2 Sample Code

```
mod1 <- glm(y~x, data=df,family="poisson")
summary(mod1) # outputs coefficients

library(emmeans)
em_mod_glm <- emmeans(mod1,~x,type="response")
summary(em_mod_lm,infer=T)

con_mod_glm <- contrast(em_mod_lm,method="pairwise",type="response")
summary(con_mod_glm,infer=T)
```