Generalized Linear Models

1 Generalized Linear Models: 3 Components

1.1 Response variable, Y

1.2 Explanatory (or predictor) variable(s)

1.3 Link function, $g()$
Links explanatory variable(s) with mean(Y)
If $\eta = \beta_0 + \beta_1 x_1$ then $\mu_{y|x} = g(\eta)$ and $\eta = g^{-1}(\mu_{y|x})$

2 To Specify Model:

2.1 Explanatory variable coefficients, β_i

2.2 Link function, $g()$

2.3 Probability distribution of Y

3 Logistic Regression

3.1 Response variable: binary [0,1]
define $\pi_i = P(Y_i = 1)$, then $P(Y_i = 0) = 1 - \pi_i$
(note: Y, π not normally distributed)

3.2 Explanatory variable(s): continuous or discrete: $x_1, x_2, \ldots, x_p$

3.3 Link function: $g(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$

$\pi/(1-\pi)$ is the "odds ratio"

inverse of link function: $\pi(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$

4 Univariate Logistic Regression Model

4.1 Model: $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1$

4.2 Odds of positive response (Y=1): $\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 x_1)$

4.3 Probability of positive response (Y=1): $\pi = \frac{\exp(\beta_0 + \beta_1 x_1)}{1+\exp(\beta_0 + \beta_1 x_1)}$

Logistic function maps $x \in [-\infty, \infty]$ onto $\pi \in [0,1]$

(continued)

5 Multivariate Logistic Regression Model

5.1 Model: $\log\left(\dfrac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$

5.2 Odds of positive response (Y=1): $\dfrac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$

5.3 Probability of positive response (Y=1): $\pi = \dfrac{\exp(\beta_0 + \beta_1 x_1) + \beta_2 x_2 + \cdots + \beta_p x_p}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}$

6 Measure of Model Fit

log-likelihood R^2: $R^2 = 1 - \dfrac{\text{residual deviance}}{\text{null deviance}}$

7 Fitting Logistic Regression Models in *R*

```
result <- glm(formula, family=..., data=...)
```

e.g.,
```
> result <- glm(response ~ x1, family=binomial)
> summary(result)
```

References

Chambers JM, Hastie TJ, eds. 1993. *Statistical Models in S*. Chapman & Hall, New York.

Gardner M. 2012. *Statistics for Ecologists Using R and Excel.* Pelagic Publishing, Exeter, UK.

Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression, 2nd ed.* John Wiley & Sons, Inc., New York.

McCullagh P, Nelder JA. 1989. *Generalized Linear Models,2nd ed.* Chapman & Hall/CRC Press, Boca Raton, FL.

Verzani J. 2005. *Using R for Introductory Statistics.* Chapman & Hall/CRC Press, Boca Raton, FL.