PLSC 503 – Spring 2022
Generalized Linear Models

April 27, 2022

"GLMs" are:

- A <u>class</u> of regression models
- ... that includes a lot of things you already know about:
  - Linear-normal regression (/ "OLS")
  - (Binary-response) logit / probit
  - Ordinal-response models
  - Nominal-response models
  - Models for event counts
  - Others...
- ...and unites them all under a common framework for estimation and inference.

The Exponential Family

Start with a random variable with density:

$$f(z|\psi) = \Pr(Z = z|\psi)$$

Exponential if:

$$f(z|\psi) = r(z)s(\psi)\exp[q(z)h(\psi)]$$

provided that $r(z) > 0$ and $s(\psi) > 0$.

$$f(z|\psi) = \exp[\underbrace{\ln r(z) + \ln s(\psi)}_{\text{“additive”}} + \underbrace{q(z)h(\psi)}_{\text{“interactive”}}]$$

Canonical Forms

"Canonical form" is when $q(z) = z$. That means we can always transform z into our usual "response" variable:

$$y = q(z)$$

$h(\theta)$ is the "natural parameter" of the distribution, often written:

$$\theta = h(\psi)$$

In canonical form, the distribution is:

$$f[y|\theta] = \exp[y\theta - b(\theta) + c(y)].$$

where"

- $b(\theta)$ is a "normalizing constant"
- $c(y)$ is a function solely of y
- $y\theta$ is a "multiplicative" term.

A Familiar Family Member: Poisson

Poisson density:

$$f(y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}.$$

Rewritten:

$$\begin{aligned} f(y|\lambda) &= \exp\left\{\ln\left[\exp(-\lambda)\lambda^y/y!\right]\right\} \\ &= \exp[\underbrace{y\ln(\lambda)}_{y\theta} - \underbrace{\lambda}_{b(\theta)} - \underbrace{\ln(y!)}_{c(y)}] \end{aligned}$$

Other parameters = "nuisance" parameters...

$$f(y|\theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right]$$

Familiar Family Member II: Normal

$$f(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{(y-\mu)^2}{2\sigma^2}\right]$$

$$\begin{aligned}
f(y|\mu,\sigma^2) &= \exp\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2)\right] \\
&= \exp\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}y^2 + \frac{1}{2\sigma^2}2y\mu - \frac{1}{2\sigma^2}\mu^2\right] \\
&= \exp\left[\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right] \\
&= \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} + \frac{-1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]\right\}
\end{aligned}$$

Normal, continued

$$f(y|\mu, \sigma^2) = \exp\left\{ \frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} + \frac{-1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right] \right\}$$

$\theta = \mu$, so:

- $y\theta = y\mu$
- $b(\theta) = \frac{\mu^2}{2}$
- $a(\phi) = \sigma^2$
- $c(y, \phi) = \frac{-1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]$

Other Family Members

- Binomial ($\supset$ Bernoulli; also Multinomial)
- Exponential
- Gamma
- Logarithmic
- Inverse Gaussian
- Negative Binomial
- others...

Little Red Likelihood

$$
\begin{aligned}
\ln L(\theta, \phi | y) &= \ln f(y|\theta, \phi) \\
&= \ln \left\{ \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \right\} \\
&= \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ln L(\theta, \phi | y)}{\partial \theta} \equiv \mathbf{S} &= \frac{\partial}{\partial \theta} \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \\
&= \frac{y - \frac{\partial}{\partial \theta} b(\theta)}{a(\phi)}.
\end{aligned}
$$

Among family members:

- $\mathbf{S}$ is a sufficient statistic for θ.
- $\mathrm{E}(\mathbf{S}) = 0$.
- $\mathrm{Var}(\mathbf{S}) \equiv \mathcal{I}(\theta) = \mathrm{E}[(\mathbf{S})^2 | \theta]$

For GLMs, under some standard regularity conditions:

$$\mathrm{E}(Y) = \frac{\partial}{\partial\theta} b(\theta)$$

and

$$\mathrm{Var}(Y) = a(\phi) \frac{\partial^2}{\partial\theta^2} b(\theta)$$

Example: Poisson Again

$$\begin{aligned}
\mathsf{E}(Y) &= \frac{\partial}{\partial\theta}\exp(\theta) \\
&= \exp(\theta)|_{\theta=\ln(\lambda)} \\
&= \lambda
\end{aligned}$$

$$\begin{aligned}
\mathsf{Var}(Y) &= 1 \times \frac{\partial^2}{\partial\theta^2}\exp(\theta)|_{\theta=\ln(\lambda)} \\
&= \exp[\ln(\lambda)] \\
&= \lambda
\end{aligned}$$

Example: Normal Again

$$
\begin{aligned}
\mathsf{E}(Y) &= \frac{\partial}{\partial\theta}\left(\frac{\theta^2}{2}\right) \\
&= \theta|_{\theta=\mu} \\
&= \mu
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{Var}(Y) &= \sigma^2 \times \frac{\partial^2}{\partial\theta^2}\left(\frac{\theta^2}{2}\right) \\
&= \sigma^2 \times \frac{\partial}{\partial\theta}\theta \\
&= \sigma^2
\end{aligned}
$$

Linear model:

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + u_i$$

$$\mathrm{E}(Y_i) \equiv \boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$$

The "generalized" part:

$$g(\boldsymbol{\mu}_i) = \mathbf{X}_i \boldsymbol{\beta}.$$

$$\begin{aligned} \boldsymbol{\eta}_i &= \mathbf{X}_i \boldsymbol{\beta} \\ &= g(\boldsymbol{\mu}_i) \end{aligned}$$

$$\begin{aligned} \boldsymbol{\mu}_i &= g^{-1}(\boldsymbol{\eta}_i) \\ &= g^{-1}(\mathbf{X}_i \boldsymbol{\beta}) \end{aligned}$$

GLMs

Random component $\sim$ EF$(\cdot)$ with

$$\mathsf{E}(Y_i) = \boldsymbol{\mu}_i.$$

Systematic component:

$$g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i$$

"Link function":

$$g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i$$

or (equivalently):

$$g^{-1}(\boldsymbol{\eta}_i) = \boldsymbol{\mu}_i.$$

The Return of The Family

$$\begin{aligned} \boldsymbol{\theta}_i &= g(\boldsymbol{\mu}_i) \\ &= \boldsymbol{\eta}_i \\ &= \mathbf{X}_i \boldsymbol{\beta} \end{aligned}$$

$$g^{-1}(\boldsymbol{\theta}_i) = \boldsymbol{\eta}_i$$

GLM Example: Linear-Normal

Density = Normal:

$$f(y|\mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2)$$

Link function = Identity:

$$\boldsymbol{\mu}_i = \boldsymbol{\eta}_i$$

Yields:

$$\begin{aligned} \boldsymbol{\mu}_i \equiv \boldsymbol{\theta}_i &= \boldsymbol{\eta}_i \\ Y_i &\sim \mathcal{N}(\boldsymbol{\mu}_i, \sigma^2) \end{aligned}$$

GLM Example: Binary

Density = Bernoulli:

$$f(y|\pi) = \pi^y (1-\pi)^{1-y}$$

Link function = Logistic:

$$\boldsymbol{\theta}_i = \ln\left(\frac{\boldsymbol{\mu}_i}{1-\boldsymbol{\mu}_i}\right)$$

Yields:

$$\begin{aligned}
\boldsymbol{\mu}_i &= g^{-1}(\boldsymbol{\theta}_i) \\
&= \frac{\exp(\boldsymbol{\eta}_i)}{1+\exp(\boldsymbol{\eta}_i)} \\
Y_i &\sim \text{Bernoulli}(\boldsymbol{\mu}_i)
\end{aligned}$$

GLM Example: Counts (Independent Events)

Density = Poisson:

$$f(y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}$$

Link function = Exponential:

$$\ln(\lambda_i) = \boldsymbol{\eta}_i$$

Yields:

$$\begin{aligned}\boldsymbol{\mu}_i &= g^{-1}(\boldsymbol{\theta}_i) \\ &= \exp(\boldsymbol{\eta}_i) \\ Y_i &\sim \text{Poisson}(\lambda_i)\end{aligned}$$

Common GLM Flavors

Distribution	Range of Y	Link(s) $g(\cdot)$	Inverse Link $g^{-1}(\cdot)$
Normal	$(-\infty, \infty)$	Identity: $\boldsymbol{\theta} = \boldsymbol{\mu}$ (Canonical)	$\boldsymbol{\theta}$
Binomial	$\{0,...n\}$	Logit: $\boldsymbol{\theta} = \ln\left(\frac{\boldsymbol{\mu}}{1-\boldsymbol{\mu}}\right)$ (Canonical)	$\frac{\exp(\boldsymbol{\theta})}{1+\exp(\boldsymbol{\theta})}$
		Probit: $\boldsymbol{\theta} = \Phi^{-1}(\boldsymbol{\mu})$	$\Phi(\boldsymbol{\theta})$
		C-Log-Log: $\boldsymbol{\theta} = \ln[-\ln(1-\boldsymbol{\mu})]$	$1 - \exp[-\exp(\boldsymbol{\theta})]$
Bernoulli	$\{0,1\}$	(same as Binomial)	(same as Binomial)
Multinomial	$\{0,...J\}$	(same as Binomial)	(same as Binomial)
Poisson	$[0, \infty]$ (integers)	Log: $\boldsymbol{\theta} = \ln(\boldsymbol{\mu})$ (Canonical)	$\exp(\boldsymbol{\theta})$
Gamma	$(0, \infty)$	Reciprocal: $\boldsymbol{\theta} = -\frac{1}{\boldsymbol{\mu}}$ (Canonical)	$-\frac{1}{\boldsymbol{\theta}}$

Note: The Bernoulli is a special case of the Binomial with $n = 1$. The multinomial is the J-outcome variant of the Binomial, and is also related to the Poisson (see, e.g., Agresti 2002).

GLMs: How-To

- Pick your (exponential family) density $f(Y)$
- Pick your link function $g(\cdot)$
- Specify your model (i.e., choose $\mathbf{X}$s)
- Estimate!

- MLE
- IRLS ($\approx$ MLE):

$$\hat{\boldsymbol{\beta}}^{(t+1)} = [\mathbf{X}'\mathbf{W}^{(t)}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}^{(t)}\mathbf{z}^{(t)}$$

with

$$\underset{N\times N}{\mathbf{W}^{(t)}} = \mathrm{diag}\left[\frac{\left(\partial\mu_i^{(t)}/\partial\eta_i^{(t)}\right)^2}{\mathrm{Var}(Y_i)}\right]$$

and

$$\mathbf{z}^{(t)} \quad = \quad \boldsymbol{\eta}^{(t)} + (Y - \boldsymbol{\mu}^{(t)})\left(\frac{\partial\boldsymbol{\eta}^{(t)}}{\partial\boldsymbol{\mu}}\right).$$

At iteration t:

1. Calculate $\mathbf{z}^{(t)}$, $\mathbf{W}^{(t)}$
2. Regress $\mathbf{z}^{(t)}$ on $\mathbf{X}$, using $\mathbf{W}^{(t)}$ as weights, to obtain $\hat{\boldsymbol{\beta}}^{(t+1)}$
3. Generate $\boldsymbol{\eta}^{(t+1)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(t+1)}$
4. Generate $\boldsymbol{\mu}^{(t+1)} = g^{-1}(\boldsymbol{\eta}^{(t+1)})$
5. Use $\boldsymbol{\eta}^{(t+1)}$ and $\boldsymbol{\mu}^{(t+1)}$ to calculate $\mathbf{z}^{(t+1)}$ and $\mathbf{W}^{(t+1)}$
6. Repeat until convergence.

Residuals

"Response" Residuals:

$$\begin{aligned}\hat{u}_i &= Y_i - \hat{\mu}_i \\ &= Y_i - g^{-1}(\mathbf{X}_i\hat{\beta})\end{aligned}$$

"Pearson" Residuals:

$$\hat{P}_i = \frac{\hat{u}_i}{[\text{Var}(\hat{u}_i)]^{1/2}}$$

More Residuals

"Deviance":

$$\begin{aligned}
\hat{d}_i &= -2[\ln L_i(\hat{\theta}) - \ln L_i(\theta_S)] \\
&= 2\left\{\left[\frac{Y_i\theta_S - b(\theta_S)}{a(\phi)} + c(Y_i,\phi)\right] - \left[\frac{Y_i\hat{\theta} - b(\hat{\theta})}{a(\phi)} + c(Y_i,\phi)\right]\right\} \\
&= 2\left[\frac{Y_i(\theta_S - \hat{\theta}) - b(\theta_S) + b(\hat{\theta})}{a(\phi)}\right]
\end{aligned}$$

"Deviance" Residuals:

$$\hat{r}_{Di} = \left(\frac{\hat{u}_i}{|\hat{u}_i|}\right)\sqrt{\hat{d}_i^2}$$

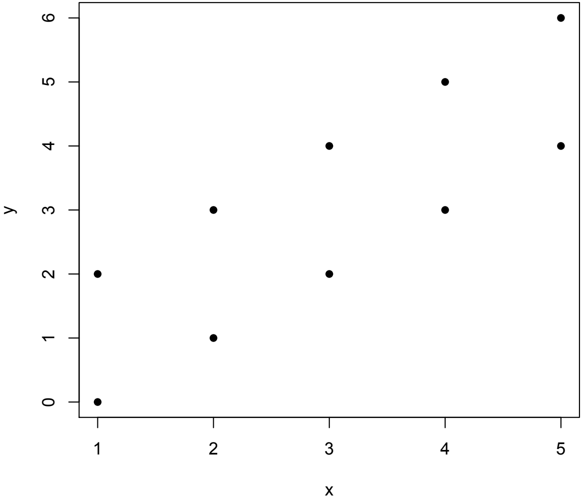
Toy Example: Linear-Normal

$$\begin{aligned} X &= \{1, 1, 2, 2, 3, 3, 4, 4, 5, 5\} \\ Y &= \{0, 2, 1, 3, 2, 4, 3, 5, 4, 6\} \end{aligned}$$

$$\begin{aligned} Y_i &= 0 + 1X_i + u_i \\ \hat{u}_i^2 &= 1 \,\forall\, i \end{aligned}$$

$$\begin{aligned} \text{“TSS”} \equiv \sum (Y_i - \bar{Y})^2 &= 30 \\ \text{“RSS”} \equiv \sum \hat{u}_i^2 &= 10 \\ \text{“MSS” / “ESS”} &= 20 \end{aligned}$$

Toy Example: Plot



Toy Example: OLS

```
> linmod<-lm(y~x)
> summary(linmod)

Call:
lm(formula = y ~ x)

Residuals:
       Min          1Q      Median          3Q         Max
-1.000e+00 -1.000e+00  1.110e-16  1.000e+00  1.000e+00

Coefficients:
              Estimate Std. Error   t value Pr(>|t|)
(Intercept) -5.617e-16  8.292e-01 -6.77e-16  1.00000
x            1.000e+00  2.500e-01         4  0.00395 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.118 on 8 degrees of freedom
Multiple R-squared: 0.6667,Adjusted R-squared: 0.625
F-statistic:    16 on 1 and 8 DF,  p-value: 0.00395
```

Toy Example: Linear-Normal GLM

```
> linglm<-glm(y~x,family="gaussian")
> summary(linglm)

Deviance Residuals:
       Min          1Q      Median          3Q         Max
-1.000e+00  -1.000e+00  -5.551e-17   1.000e+00   1.000e+00

Coefficients:
              Estimate Std. Error   t value Pr(>|t|)
(Intercept) -5.617e-16  8.292e-01 -6.77e-16  1.00000
x            1.000e+00  2.500e-01         4  0.00395 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for gaussian family taken to be 1.25)

    Null deviance: 30  on 9  degrees of freedom
Residual deviance: 10  on 8  degrees of freedom
AIC: 34.379

Number of Fisher Scoring iterations: 2
```

Better GLM Example: Political Knowledge

- 2008 NES political knowledge
- Identify Speaker of the House, VP, British PM, and Chief Justice
- $Y_i =$ number of correct answers (out of four)

$$f(Y_i, p_i) = \binom{4}{Y_i} p_i^{Y_i} (1 - p_i)^{4 - Y_i}$$

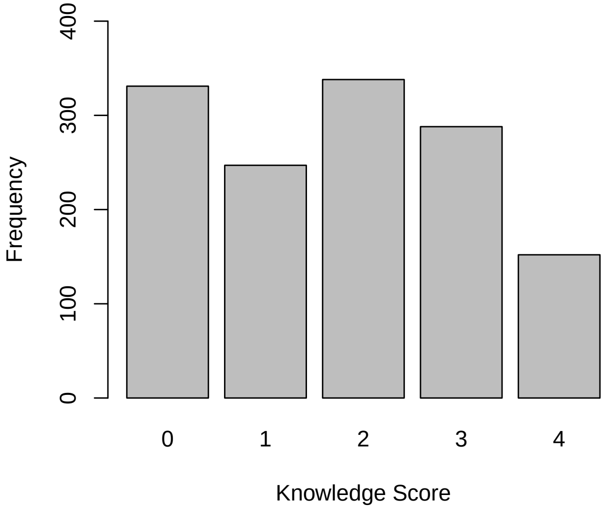
$$Y \sim \text{Binomial}(4, p)$$

$$\text{E}(Y_i) = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta})}$$

GLM Example Data (2008 NES)

```
> summary(NES08[,4:16])

   knowledge          sex             race             age            female     
 Min.   :0.00   Min.   :1.00   Min.   :1.00   Min.   :18.0   Min.   :0.000  
 1st Qu.:1.00   1st Qu.:1.00   1st Qu.:1.00   1st Qu.:33.0   1st Qu.:0.000  
 Median :2.00   Median :2.00   Median :1.00   Median :45.0   Median :1.000  
 Mean   :1.77   Mean   :1.53   Mean   :1.56   Mean   :46.2   Mean   :0.535  
 3rd Qu.:3.00   3rd Qu.:2.00   3rd Qu.:2.00   3rd Qu.:57.0   3rd Qu.:1.000  
 Max.   :4.00   Max.   :2.00   Max.   :7.00   Max.   :90.0   Max.   :1.000  

     white          oftenvote   conservative      prayfreq     heterosexual  
 Min.   :0.000   Min.   :1.0   Min.   :1.00   Min.   :1.00   Min.   :0.000  
 1st Qu.:0.000   1st Qu.:2.0   1st Qu.:3.00   1st Qu.:3.00   1st Qu.:1.000  
 Median :1.000   Median :3.0   Median :4.00   Median :4.00   Median :1.000  
 Mean   :0.692   Mean   :2.8   Mean   :4.15   Mean   :3.55   Mean   :0.954  
 3rd Qu.:1.000   3rd Qu.:4.0   3rd Qu.:5.00   3rd Qu.:5.00   3rd Qu.:1.000  
 Max.   :1.000   Max.   :4.0   Max.   :7.00   Max.   :5.00   Max.   :1.000  

    married        yrsofschool        income    
 Min.   :0.000   Min.   : 0.0   Min.   : 1.0  
 1st Qu.:0.000   1st Qu.:12.0   1st Qu.: 7.0  
 Median :0.000   Median :14.0   Median :12.0  
 Mean   :0.451   Mean   :13.6   Mean   :11.4  
 3rd Qu.:1.000   3rd Qu.:16.0   3rd Qu.:16.0  
 Max.   :1.000   Max.   :17.0   Max.   :25.0  
```

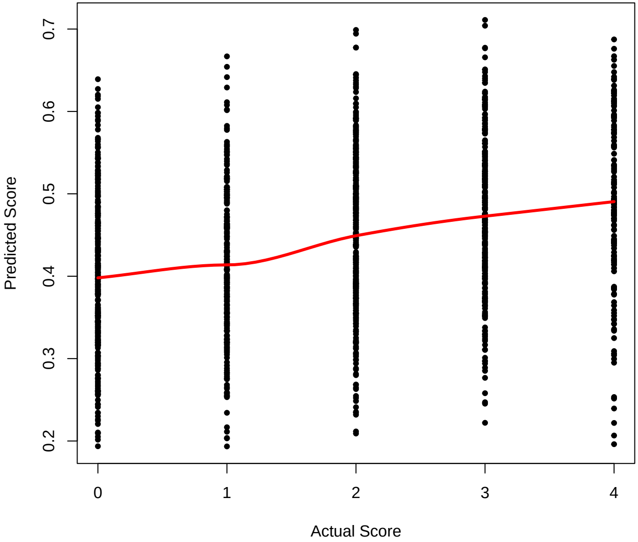
Political Knowledge (2008 NES)



GLM Results

```
> nes08.binom<-glm(cbind(knowledge,4-knowledge)~age+female+white+oftenvote+conservative+
+                     prayfreq+heterosexual+married+yrsofschool+income,data=NES08,family=binomial)
> summary(nes08.binom)

Call:
glm(formula = cbind(knowledge, 4 - knowledge) ~ age + female +
    white + oftenvote + conservative + prayfreq + heterosexual +
    married + yrsofschool + income, family = binomial, data = NES08)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.856  -1.339  -0.034   1.015   3.610

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.12090    0.24933   -8.51  < 2e-16 ***
age           0.01097    0.00192    5.72  1.0e-08 ***
female       -0.20793    0.05961   -3.49  0.00049 ***
white         0.15193    0.06487    2.34  0.01917 *
oftenvote     0.09633    0.02755    3.50  0.00047 ***
conservative -0.02160    0.01935   -1.12  0.26425
prayfreq     -0.05034    0.02227   -2.26  0.02378 *
heterosexual -0.06604    0.13853   -0.48  0.63358
married       0.16896    0.05848    2.89  0.00386 **
yrsofschool   0.09320    0.01316    7.08  1.4e-12 ***
income        0.00831    0.00527    1.58  0.11515
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3163.9  on 1355  degrees of freedom
Residual deviance: 2933.8  on 1345  degrees of freedom
AIC: 4542

Number of Fisher Scoring iterations: 4
```

Example: In-Sample Predictions



GLMs: Other Topics + Extensions

Other Topics:

- Residual-Based Diagnostics (leverage, model fit, etc.)
- Generalizations for Overdispersion (binomial)
- Joint Mean-Dispersion Models

Extensions:

- Bias-reduced models (a la Firth 1993)
- "Generalized additive models" (GAMs)
- "Generalized estimating equations" (GEEs)
- "Vector" GLMs (Yee and Wild 1996; Yee and Hastie 2003)

GLMs: References

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd Ed. London: Chapman & Hall.

Dobson, Annette J., and and Adrian G. Barnett. 2018. *An Introduction to Generalized Linear Models*, 4th Ed. London: Chapman & Hall.

Faraway, Julian J. 2016. *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models*, 2nd Ed. London: Chapman & Hall / CRC.

Dunn, Peter K., and Gordon K. Smyth. 2018. *Generalized Linear Models With Examples in R*. New York: Springer.

Paul Roback and Julie Legler. 2020. *Beyond Multiple Linear Regression: Applied Generalized Linear Models And Multilevel Models in R*. London: Chapman & Hall/CRC Press.