BIOSTAT 651 SUMMARY

(1) Assumptions of LM and GLM

	LM	GLM
f_{Y_i}	Normal	Exponential family
$E(Y_i)$	$x_i^T\beta$	$g^{-1}(x_i^T\beta)$
$Var(Y_i)$	σ	σ_i

(2) MLE

(a) Likelihood, score, information

$$I(\theta_0) = nI_1(\theta_0) \xleftarrow{P} J(\hat{\theta})$$

(b) Asymptotic properties

$$\hat{\theta} \sim N(\theta_0, I(\theta_0)^{-1})$$
$$U(\theta_0) \sim N(0, I(\theta_0))$$

(c) Hypothesis test for θ:
H_0: $\theta_1 = \theta_{H1}$, H_a: $\theta_1 \neq \theta_{H1}$, ($\theta^T = (\theta_1^T, \theta_2^T)$, $\mathrm{df}(\theta) = q = q_1 + q_2$)

(i) Wald test

$$(\hat{\theta}_1 - \theta_{1H})^T \hat{V}(\hat{\theta}_1)^{-1}(\hat{\theta}_1 - \theta_{1H}) \sim \chi^2_{q_1}$$

Where $\hat{V}(\hat{\theta}_1)$ is (I think) the first q_1 diagonal block matrix of $I(\hat{\theta})$.

(ii) Score test

$$U(\hat{\theta}_H)^T I(\hat{\theta}_H)^{-1} U(\hat{\theta}_H) \sim \chi^2_{q_1}$$

We use χ^2_{q1} rather than χ^2_q, even though $U(\hat{\theta}_H)$ has q elements, because the last q_2 elements of $U(\hat{\theta}_H)$ are always zero.

(iii) Likelihood ratio test

$$-2(\ell(\hat{\theta}_H) - \ell(\hat{\theta})) \sim \chi^2_{q_1}$$

(3) Exponential family:

(a) Single-parameter:

$$f(y; \theta, \phi) = \exp\left\{\frac{t(Y)\theta - b(\theta)}{a(\phi)} + c(Y, \phi)\right\}$$

where $a(\phi) > 0$, and we like the canonical form $t(Y) = Y$.

(b) Properties:

$$E(Y_i) = \mu_i = b'(\theta_i)$$
$$Var(Y_i) = b''(\theta_i)a(\phi) = v(\mu_i)a(\phi)$$

(c) Multi-parameter

$$f(y;\theta) = \exp\left\{\sum_{j=1}^{m} t_j(Y)\theta_j - b(\theta) + c(Y)\right\}$$

(4) Linking function

$$g(\mu_i) = \eta_i = x_i^T\beta$$

(a) Canonical link ($\theta_i = \eta_i$):

$$v(\mu_i) = b''(\theta_i) = \frac{\partial \mu_i}{\partial \theta_i} = \frac{1}{g'(\mu_i)}$$

for common distributions:

Distribution	$g(\mu_i)$	$v(\mu_i)$	θ	$b(\theta)$	$a(\phi)$
Normal	μ_i	1	μ	$\frac{1}{2}\theta^2$	σ^2
Bernoulli	$\log\{\frac{\mu_i}{1-\mu_i}\}$	$\mu_i(1-\mu_i)$	$\log\{\frac{\pi}{1-\pi}\}$	$\log(e^\theta+1)$	1
Poisson	$\log(\mu_i)$	μ_i	$\log(\lambda)$	e^θ	1

(5) Point estimation

$$U(\beta) = X^T[a(\phi)\Delta V]^{-1}(Y-\mu)$$
$$I(\beta) = X^T[a(\phi)\Delta V\Delta]^{-1}X$$

where

$$V = \text{diag}[v(\mu_i)] = \text{diag}[b''(\theta_i)]$$
$$\Delta = \text{diag}[g'(\mu_i)]$$

(a) Canonical link:

$$U(\beta) = X^T[a(\phi)]^{-1}(Y-\mu)$$
$$J(\beta) = I(\beta) = X^T[a(\phi)\Delta]^{-1}X$$

since $\Delta V = I$. And find $\hat\beta$ by solving

$$X^T(Y-\mu) = 0$$

(6) Computation methods

(a) Newton-Raphson

$$\hat\beta_{j+1} = \hat\beta_j + J(\hat\beta_j)^{-1}U(\hat\beta_j)$$

(b) Fisher scoring

$$\hat\beta_{j+1} = \hat\beta_j + I(\hat\beta_j)^{-1}U(\hat\beta_j)$$

More robust, convergence takes longer, same as NR if canonical link

(c) IRWLS

$$\hat\beta_{j+1} = (X^TV_jX)^{-1}X^TV_jZ_j$$
$$Z_j = \eta_j + V_j^{-1}(Y-\mu_j)$$

This is for canonical link. For non-canonical, switch to the appropriate $U(\beta)$ and $I(\beta)$.

(7) Hypothesis test for β

(a) Wald test (H_0: $C\beta - d = 0$)

$$(C\beta - d)^T (CI(\hat{\beta})^{-1} C^T)^{-1} (C\beta - d) \sim X_r^2$$

where $r = \text{rank}(C)$.

(b) Score test (H_0: $\beta_1 - d = 0$)

$$U(\hat{\beta}_H)^T I(\hat{\beta}_H)^{-1} U(\hat{\beta}_H) \sim \chi^2_{q_1}$$

(c) Likelihood ratio test (H_0: $\beta_1 - d = 0$)

$$-2(\ell(\hat{\beta}_H) - \ell(\hat{\beta})) \sim \chi^2_{q_1}$$

(8) Model diagnostics

(a) Goodness of fit tests

(i) Deviance

$$D = 2\sum\{Y_i(\tilde{\theta}_i - \hat{\theta}_i) - [b(\tilde{\theta}_i) - b(\hat{\theta}_i)]\} = \sum[\hat{r}_i^D]^2$$

$$D^* = 2[\ell(\tilde{\theta}) - \ell(\hat{\theta})] = D/a(\phi) \sim \chi^2_{n-q} \quad \text{when model fits well}$$

Notice that $D_0^* - D_1^* = X^2_{LRT}$ is the likelihood ratio statistic.

(ii) Pearson chi-square statistic

$$X_P^2 = \sum(Y_i - \hat{\mu}_i)^2/\hat{V}(Y_i) = \sum[\hat{r}_i^P]^2 \sim \chi^2_{n-q} \quad \text{when model fits well}$$

(b) Leverage: h_{ii} the diagonal elements of the hat matrix

$$H = V^{1/2}X(X^TVX)^{-1}X^TV^{1/2}$$

Then the standarized residuals are

$$\hat{r}_i^{DS} = \hat{r}_i^D/\sqrt{1 - h_{ii}}$$
$$\hat{r}_i^{PS} = \hat{r}_i^P/\sqrt{1 - h_{ii}}$$

(c) Influence: Cook's distance

$$D_i = \frac{1}{q}\frac{h_{ii}}{1 - h_{ii}}(r_i^{PS})^2$$

(d) Multicollinearity: Variance inflation factor

$$VIF_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is obtained by regressing the j^{th} variable against all the others in the weighted predictor matrix $V^{1/2}X$.