Today

- deviance and scaled deviance — SM errata
- nonlinear least squares — §10.2
- proportion data and contingency tables — §10.4, 10.5.2
- CD on observational studies
- HW 2 due February 28
- You should read: §10.1, 10.2 (excl. Ex. 10.7), 10.3, 10.4, 10.5.2 (excl. Ex.10.22, p.505–511), 10.6
- by end of reading week

Deviance and scaled deviance

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$ SM §10.2
- assume ϕ is known
- log-likelihood function $\ell(\eta; y) = \sum_{j=1}^n \log f(y_j; \eta_j, \phi)$
- $\tilde\eta = \arg\sup_\eta \ell(\eta; y)$
- now suppose $\eta_j = \eta_j(\beta), \beta \in \mathbb{R}^p$
- $\hat\eta = \eta(\hat\beta)$
- "scaled deviance"

$$D = 2 \sum \{\log f(y_j; \tilde\eta_j) - \log f(y_j; \hat\eta_j)\}$$

- example: $y_j \sim N(\eta_j; \sigma^2)$; $\quad \tilde\eta_j = y_j$; $\quad D = \sum(y_j - \hat\eta_j)^2/\sigma^2$

Deviance and scaled deviance

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \dots, n$ SM §10.2
- assume ϕ is known
- log-likelihood function $\ell(\eta; y) = \sum_{j=1}^{n} \log f(y_j; \eta_j, \phi)$
- $\tilde{\eta} = \arg\sup_{\eta} \ell(\eta; y)$
- now suppose $\eta_j = \eta_j(\beta), \beta \in \mathbb{R}^p$
- $\hat{\eta} = \eta(\hat{\beta})$
- "scaled deviance"

$$D = 2 \sum \{\log f(y_j; \tilde{\eta}_j) - \log f(y_j; \hat{\eta}_j)\}$$

- example: $y_j \sim N(\eta_j; \sigma^2)$; $\quad \tilde{\eta}_j = y_j$; $\quad D = \sum (y_j - \hat{\eta}_j)^2 / \sigma^2$

Deviance and scaled deviance

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$ SM §10.2
- assume ϕ is known
- log-likelihood function $\ell(\eta; y) = \sum_{j=1}^{n} \log f(y_j; \eta_j, \phi)$
- $\tilde{\eta} = \arg\sup_\eta \ell(\eta; y)$
- now suppose $\eta_j = \eta_j(\beta), \beta \in \mathbb{R}^p$
- $\hat{\eta} = \eta(\hat{\beta})$
- "scaled deviance"

$$D = 2\sum\{\log f(y_j; \tilde{\eta}_j) - \log f(y_j; \hat{\eta}_j)\}$$

- example: $y_j \sim N(\eta_j; \sigma^2)$; $\quad \tilde{\eta}_j = y_j$; $\quad D = \sum(y_j - \hat{\eta}_j)^2/\sigma^2$

Deviance and scaled deviance

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \dots, n$ SM §10.2
- assume ϕ is known
- log-likelihood function $\ell(\eta; y) = \sum_{j=1}^n \log f(y_j; \eta_j, \phi)$
- $\tilde\eta = \arg\sup_\eta \ell(\eta; y)$

- now suppose $\eta_j = \eta_j(\beta)$, $\beta \in \mathbb{R}^p$
- $\hat\eta = \eta(\hat\beta)$
- "scaled deviance"

$$D = 2 \sum \{\log f(y_j; \tilde\eta_j) - \log f(y_j; \hat\eta_j)\}$$

- example: $y_j \sim N(\eta_j; \sigma^2)$; $\tilde\eta_j = y_j$; $D = \sum (y_j - \hat\eta_j)^2 / \sigma^2$

Deviance and scaled deviance

SM §10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- assume ϕ is known
- log-likelihood function $\ell(\eta; y) = \sum_{j=1}^{n} \log f(y_j; \eta_j, \phi)$
- $\tilde\eta = \arg\sup_\eta \ell(\eta; y)$

- now suppose $\eta_j = \eta_j(\beta), \beta \in \mathbb{R}^p$
- $\hat\eta = \eta(\hat\beta)$
- "scaled deviance"

$$D = 2\sum\{\log f(y_j; \tilde\eta_j) - \log f(y_j; \hat\eta_j)\}$$

- example: $y_j \sim N(\eta_j; \sigma^2)$; $\quad \tilde\eta_j = y_j$; $\quad D = \sum(y_j - \hat\eta_j)^2/\sigma^2$

Deviance and scaled deviance

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$ SM §10.2
- assume ϕ is known
- log-likelihood function $\ell(\eta; y) = \sum_{j=1}^{n} \log f(y_j; \eta_j, \phi)$
- $\tilde{\eta} = \arg\sup_{\eta} \ell(\eta; y)$

- now suppose $\eta_j = \eta_j(\beta), \beta \in \mathbb{R}^p$
- $\hat{\eta} = \eta(\hat{\beta})$
- "scaled deviance"

$$D = 2 \sum \{\log f(y_j; \tilde{\eta}_j) - \log f(y_j; \hat{\eta}_j)\}$$

- example: $y_j \sim N(\eta_j; \sigma^2)$; $\tilde{\eta}_j = y_j$; $D = \sum (y_j - \hat{\eta}_j)^2 / \sigma^2$

Deviance and scaled deviance

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$ SM §10.2
- assume ϕ is known
- log-likelihood function $\ell(\eta; y) = \sum_{j=1}^{n} \log f(y_j; \eta_j, \phi)$
- $\tilde{\eta} = \arg\sup_{\eta} \ell(\eta; y)$

- now suppose $\eta_j = \eta_j(\beta), \beta \in \mathbb{R}^p$
- $\hat{\eta} = \eta(\hat{\beta})$
- "scaled deviance"

$$D = 2 \sum \{\log f(y_j; \tilde{\eta}_j) - \log f(y_j; \hat{\eta}_j)\}$$

- example: $y_j \sim N(\eta_j; \sigma^2)$; $\tilde{\eta}_j = y_j$; $D = \sum (y_j - \hat{\eta}_j)^2 / \sigma^2$

Deviance and scaled deviance

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$ SM §10.2
- assume ϕ is known
- log-likelihood function $\ell(\eta; y) = \sum_{j=1}^{n} \log f(y_j; \eta_j, \phi)$
- $\tilde{\eta} = \arg\sup_\eta \ell(\eta; y)$

- now suppose $\eta_j = \eta_j(\beta), \beta \in \mathbb{R}^p$
- $\hat{\eta} = \eta(\hat{\beta})$
- "scaled deviance"

$$D = 2 \sum \{\log f(y_j; \tilde{\eta}_j) - \log f(y_j; \hat{\eta}_j)\}$$

- example: $y_j \sim N(\eta_j; \sigma^2); \quad \tilde{\eta}_j = y_j; \quad D = \sum (y_j - \hat{\eta}_j)^2 / \sigma^2$

... deviance and scaled deviance

- if two nested models are to be compared, $A \subset B$
- then two scaled deviances are computed
- $D_A = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^A)\}$
- $D_B = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^B)\}$
- difference is free of $\tilde{\eta}$
- difference is log-likelihood ratio statistic for testing model A (within model B)
- $D_A - D_B \dot{\sim} \chi^2_{p_B - p_A}$
- if ϕ is unknown, then must be estimated, before D_A or D_B can be computed

... deviance and scaled deviance

- if two nested models are to be compared, $A \subset B$
- then two scaled deviances are computed
- $D_A = 2\sum\{\log f(y_j; \tilde\eta_j) - \log f(y_j; \hat\eta_j^A)\}$
- $D_B = 2\sum\{\log f(y_j; \tilde\eta_j) - \log f(y_j; \hat\eta_j^B)\}$
- difference is free of $\tilde\eta$
- difference is log-likelihood ratio statistic for testing model A (within model B)
- $D_A - D_B \dot\sim \chi^2_{p_B - p_A}$
- if ϕ is unknown, then must be estimated, before D_A or D_B can be computed

... deviance and scaled deviance

- if two nested models are to be compared, $A \subset B$
- then two scaled deviances are computed
- $D_A = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^A)\}$
- $D_B = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^B)\}$
- difference is free of $\tilde{\eta}$
- difference is log-likelihood ratio statistic for testing model A (within model B)
- $D_A - D_B \dot{\sim} \chi^2_{p_B - p_A}$
- if ϕ is unknown, then must be estimated, before D_A or D_B can be computed

... deviance and scaled deviance

- if two nested models are to be compared, $A \subset B$
- then two scaled deviances are computed
- $D_A = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^A)\}$
- $D_B = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^B)\}$
- difference is free of $\tilde{\eta}$
- difference is log-likelihood ratio statistic for testing model A (within model B)
- $D_A - D_B \dot{\sim} \chi^2_{p_B - p_A}$
- if ϕ is unknown, then must be estimated, before D_A or D_B can be computed

... deviance and scaled deviance

- if two nested models are to be compared, $A \subset B$
- then two scaled deviances are computed
- $D_A = 2\sum\{\log f(y_j;\tilde\eta_j) - \log f(y_j;\hat\eta_j^A)\}$
- $D_B = 2\sum\{\log f(y_j;\tilde\eta_j) - \log f(y_j;\hat\eta_j^B)\}$
- difference is free of $\tilde\eta$
- difference is log-likelihood ratio statistic for testing model A (within model B)
- $D_A - D_B \dot\sim \chi^2_{p_B - p_A}$
- if ϕ is unknown, then must be estimated, before D_A or D_B can be computed

... deviance and scaled deviance

- if two nested models are to be compared, $A \subset B$
- then two scaled deviances are computed
- $D_A = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^A)\}$
- $D_B = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^B)\}$
- difference is free of $\tilde{\eta}$

- difference is log-likelihood ratio statistic for testing model A (within model B)
- $D_A - D_B \dot{\sim} \chi^2_{p_B - p_A}$
- if ϕ is unknown, then must be estimated, before D_A or D_B can be computed

... deviance and scaled deviance

- if two nested models are to be compared, $A \subset B$
- then two scaled deviances are computed
- $D_A = 2\sum\{\log f(y_j; \tilde{\eta}_j) - \log f(y_j; \hat{\eta}_j^A)\}$
- $D_B = 2\sum\{\log f(y_j; \tilde{\eta}_j) - \log f(y_j; \hat{\eta}_j^B)\}$
- difference is free of $\tilde{\eta}$

- difference is log-likelihood ratio statistic for testing model A (within model B)
- $D_A - D_B \dot{\sim} \chi^2_{p_B - p_A}$
- if ϕ is unknown, then must be estimated, before D_A or D_B can be computed

... deviance and scaled deviance

- if two nested models are to be compared, $A \subset B$
- then two scaled deviances are computed
- $D_A = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^A)\}$
- $D_B = 2\sum\{\log f(y_j;\tilde{\eta}_j) - \log f(y_j;\hat{\eta}_j^B)\}$
- difference is free of $\tilde{\eta}$

- difference is log-likelihood ratio statistic for testing model A (within model B)
- $D_A - D_B \dot{\sim} \chi^2_{p_B - p_A}$
- if ϕ is unknown, then must be estimated, before D_A or D_B can be computed

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $$\hat{\phi} = \frac{1}{n-p}\sum_{j=1}^{n}\frac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of var$(\hat{\beta}_j)$ by $\hat{\phi}$

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $\hat{\phi} = \dfrac{1}{n-p} \sum_{j=1}^{n} \dfrac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of $\text{var}(\hat{\beta}_j)$ by $\hat{\phi}$

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $\hat{\phi} = \dfrac{1}{n-p} \sum_{j=1}^{n} \dfrac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of $\mathrm{var}(\hat{\beta}_j)$ by $\hat{\phi}$

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $\hat{\phi} = \dfrac{1}{n-p} \sum_{j=1}^{n} \dfrac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of var$(\hat{\beta}_j)$ by $\hat{\phi}$

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $\hat{\phi} = \dfrac{1}{n-p} \sum_{j=1}^{n} \dfrac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of var($\hat{\beta}_j$) by $\hat{\phi}$

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $\hat{\phi} = \dfrac{1}{n-p} \sum_{j=1}^{n} \dfrac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of var$(\hat{\beta}_j)$ by $\hat{\phi}$

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of var$(\hat{\beta}_j)$ by $\hat{\phi}$

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $\hat{\phi} = \dfrac{1}{n-p} \sum_{j=1}^{n} \dfrac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of var$(\hat{\beta}_j)$ by $\hat{\phi}$

... deviance and scaled deviance, GLMS

- ϕ is either 1 (Binomial, Poisson) or unknown (Normal, Gamma, inverse Gaussian)
- when known, D measures 'goodness of fit' of the fitted model
- because it compares it to the saturated model $\ell(\tilde{\eta}; y)$
- when unknown, it appears as a scale factor in D
- thus $\dfrac{(D_A - D_B)/(p_B - p_A)}{\hat{\phi}} \dot{\sim} F_{p_B - p_A, p_B}$
- $\hat{\phi}$ estimated under the larger model, B
- $\hat{\phi} = \dfrac{1}{n-p} \displaystyle\sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$
- if Poisson or Binomial have over dispersion, $\hat{\phi} > 1$, then D no longer measures goodness of fit
- multiply estimate of var$(\hat{\beta}_j)$ by $\hat{\phi}$

Nonlinear models

§10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat{\beta}$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
- $$\hat{\beta} = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \qquad (1)$$
- $X = X(\hat{\beta}) = \left.\dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}}\right|_{\hat{\beta}}$
- $W = W(\hat{\beta}) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat{\beta}) = (X\beta + W^{-1} u); \quad u_j(\hat{\beta}) = \partial \ell_j(\eta) / \partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Nonlinear models

§10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat\beta$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
- $$\hat\beta = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \qquad (1)$$
- $X = X(\hat\beta) = \left. \dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}} \right|_{\hat\beta}$
- $W = W(\hat\beta) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat\beta) = (X\beta + W^{-1} u); \quad u_j(\hat\beta) = \partial \ell_j(\eta)/\partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Nonlinear models

§10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat\beta$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
-
$$\hat\beta = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \qquad (1)$$
- $X = X(\hat\beta) = \left.\dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}}\right|_{\hat\beta}$
- $W = W(\hat\beta) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat\beta) = (X\beta + W^{-1}u); \quad u_j(\hat\beta) = \partial \ell_j(\eta)/\partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Nonlinear models

§10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat\beta$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
- $$\hat\beta = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \qquad (1)$$
- $X = X(\hat\beta) = \left.\dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}}\right|_{\hat\beta}$
- $W = W(\hat\beta) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat\beta) = (X\beta + W^{-1}u); \quad u_j(\hat\beta) = \partial \ell_j(\eta)/\partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Nonlinear models

§10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat{\beta}$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
- $$\hat{\beta} = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \quad (1)$$
- $X = X(\hat{\beta}) = \left.\dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}}\right|_{\hat{\beta}}$
- $W = W(\hat{\beta}) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat{\beta}) = (X\beta + W^{-1} u); \quad u_j(\hat{\beta}) = \partial \ell_j(\eta) / \partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Nonlinear models

§10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat\beta$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
-
$$\hat\beta = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \qquad (1)$$
- $X = X(\hat\beta) = \left.\dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}}\right|_{\hat\beta}$
- $W = W(\hat\beta) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat\beta) = (X\beta + W^{-1} u); \quad u_j(\hat\beta) = \partial \ell_j(\eta) / \partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Nonlinear models §10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat{\beta}$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
- $$\hat{\beta} = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \qquad (1)$$
- $X = X(\hat{\beta}) = \left.\dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}}\right|_{\hat{\beta}}$
- $W = W(\hat{\beta}) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat{\beta}) = (X\beta + W^{-1}u); \quad u_j(\hat{\beta}) = \partial \ell_j(\eta)/\partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Nonlinear models

§10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat{\beta}$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
- $$\hat{\beta} = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \qquad (1)$$
- $X = X(\hat{\beta}) = \left.\dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}}\right|_{\hat{\beta}}$
- $W = W(\hat{\beta}) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat{\beta}) = (X\beta + W^{-1} u); \quad u_j(\hat{\beta}) = \partial \ell_j(\eta) / \partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Nonlinear models §10.2

- $y_j \sim f(y_j; \eta_j, \phi)$, independently, $j = 1, \ldots, n$
- $\eta_j = \eta_j(\beta)$
- $\hat\beta$ can be computed by iteratively re-weighted LS
- when (i) ϕ is fixed and
  (ii) $J \leftarrow I$ observed Fisher info. replaced by expected Fisher info.
- $$\hat\beta = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W z \qquad (1)$$
- $X = X(\hat\beta) = \left.\dfrac{\partial \eta(\beta)}{\partial \beta^{\mathrm{T}}}\right|_{\hat\beta}$
- $W = W(\hat\beta) = \mathrm{diag}(w_j); \quad w_j = \mathrm{E}(-\partial^2 \ell_j / \partial \eta_j^2)$
- $z = z(\hat\beta) = (X\beta + W^{-1}u); \quad u_j(\hat\beta) = \partial \ell_j(\eta)/\partial \eta_j$
- form of (1) gives definitions of residuals, influence, leverage, by analogy with linear model §8.6.1, 8.6.3, 10.2.3

Calcium data: Example 10.1

10.1 · Introduction

Table 10.1 Calcium uptake (nmoles/mg) of cells suspended in a solution of radioactive calcium, as a function of time suspended (minutes) (Rawlings, 1988, p. 403).

Time (minutes)	Calcium uptake (nmoles/mg)		
0.45	0.34170	−0.00438	0.82531
1.30	1.77967	0.95384	0.64080
2.40	1.75136	1.27497	1.17332
4.00	3.12273	2.60958	2.57429
6.10	3.17881	3.00782	2.67061
8.05	3.05959	3.94321	3.43726
11.15	4.80735	3.35583	2.78309
13.15	5.13825	4.70274	4.25702
15.00	3.60407	4.15029	3.42484

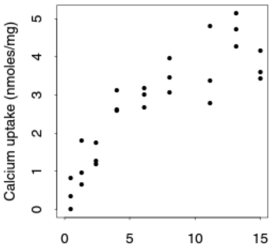


Figure 10.1 Calcium uptake (nmoles/mg) of cells suspended in a solution of radioactive calcium, as a function of time suspended (minutes).

... calcium data

- model
  $E(y_j) = \beta_0\{1 - \exp(-x_j/\beta_1)\}, \quad y_j = E(y_j) + \epsilon_j,\ \epsilon_j \sim N(0, \sigma^2)$
- fitting:

$$\min_{\beta_0, \beta_1} \sum_{j=1}^{n} (y_j - \eta_j)^2$$

- use `nls` or `nlm`; requires starting values
-

```
> library(SMPracticals); data(calcium)
> fit = nls(cal ~ b0*(1-exp(-time/b1)), data = calcium, start = list(b0=5,b1=5))
> summary(fit)
Formula: cal ~ b0 * (1 - exp(-time/b1))

Parameters:
   Estimate Std. Error t value Pr(>|t|)
b0   4.3094     0.3029  14.226 1.73e-13 ***
b1   4.7967     0.9047   5.302 1.71e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5464 on 25 degrees of freedom

Number of iterations to convergence: 3
Achieved convergence tolerance: 9.55e-07
```

... calcium data

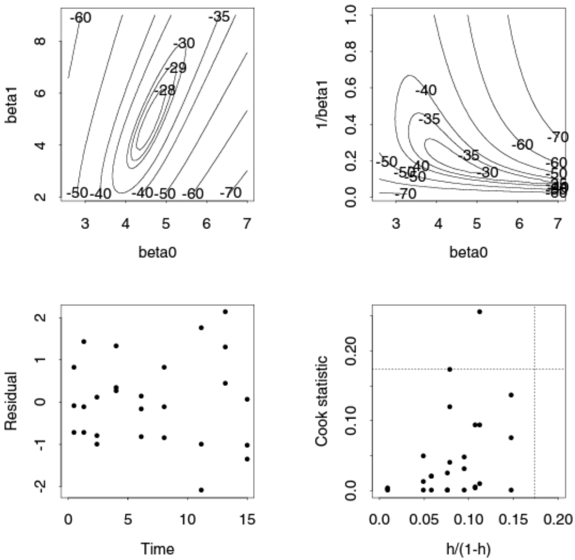


Figure 10.4 Fit of a nonlinear model to the calcium data. Upper left: contours for $\ell_p(\beta_0, \beta_1)$. Upper right: contours for $\ell_p(\beta_0, \gamma_1)$, where $\gamma_1 = 1/\beta_1$. Lower left: standardized residuals plotted against time. Lower right: plot of Cook statistics against $h/(1-h)$, where h is leverage.

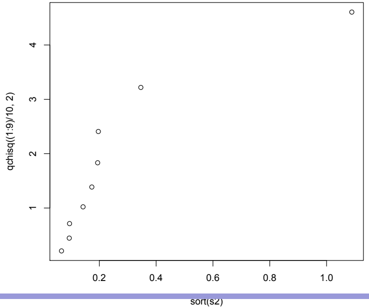
... calcium data

- there are 3 observations at each time point
- can fit a model with a different parameter for each time: $E(y_j) = \eta_j + \epsilon_j$
- the nonlinear model is nested within this; constrains η_j as above
- `anova(lm(cal ~ factor(time), data = calcium))`
- ```
  Analysis of Variance Table

  Response: cal
                Df Sum Sq Mean Sq F value    Pr(>F)
  factor(time)   8 48.437  6.0546  22.720 6.688e-08 ***
  Residuals     18  4.797  0.2665
  ```
- ```
  > deviance(fit) # 7.464514 (mistake in Davison)
  > sum(residuals(fit)^2) # 7.464514
  > (7.464514 - 4.797)/(25 - 18) # 0.3811
  > .3811/.2665
  [1] 1.429919  ## Davison has 1.53
  > pf(1.430,7,18, lower=F)
  [1] 0.2538313
  ```

... calcium data

- checking constant variance assumption
- estimates of σ^2 at each time, each with 2 degrees of freedom
- ```
  > s2 = tapply(calcium$cal, factor(calcium$time), var)
  > s2
  > s2
        0.45        1.3        2.4          4        6.1       8.05
  0.17367258 0.34616902 0.09523507 0.09422579 0.06686923 0.19656739
       11.15      13.15         15
  1.08876166 0.19415027 0.14279290
  > plot(sort(s2),qchisq((1:9)/10,2))
  ```


  ```

Example

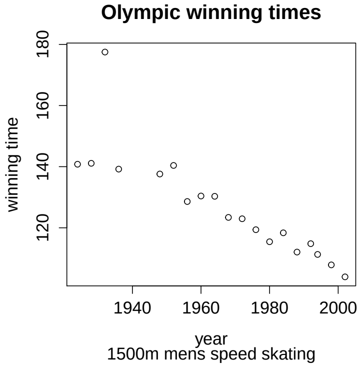
Factmonster

1500 meters

Year		Time	
1924	Clas Thunberg, FIN	2:20.8	
1928	Clas Thunberg, FIN	2:21.1	
1932	Jack Shea, USA	2:57.5	
1936	Charles Mathisen, NOR	2:19.2	**OR**
1948	Sverre Farstad, NOR	2:17.6	**OR**
1952	Hjalmar Andersen, NOR	2:20.4	
1956	(TIE) Yevgeny Grishin, USSR	2:08.6	**WR**
	& Yuri Mikhailov, USSR	2:08.6	**WR**
1960	(TIE) Roald Aas, NOR	2:10.4	
	& Yevgeny Grishin, USSR	2:10.4	
1964	Ants Antson, USSR	2:10.3	
1968	Kees Verkerk, NED	2:03.4	**OR**
1972	Ard Schenk, NED	2:02.96	**OR**
1976	Jan Egil Storholt, NOR	1:59.38	**OR**
1980	Eric Heiden, USA	1:55.44	**OR**
1984	Gaétan Boucher, CAN	1:58.36	
1988	Andre Hoffman, E. Ger	1:52.06	**WR**
1992	Johann Olav Koss, NOR	1:54.81	
1994	Johann Olav Koss, NOR	1:51.29	**WR**
1998	Aadne Sondral, NOR	1:47.87	**WR**
2002	Derek Parra, USA	1:43.95	**WR**

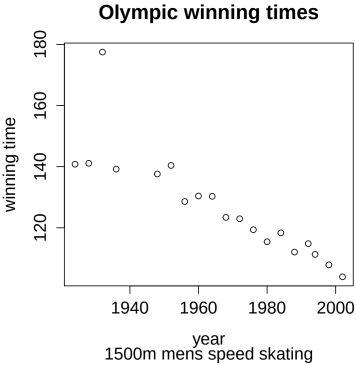
Nonlinear model

- $E(y_j) = \beta_0 + \beta_1 \exp(-\beta_2 x_j)$ "Analysis of running times in Olympic games"
- winning times are decreasing; rate of improvement is decreasing; limiting time (β_0)



Nonlinear model

- $E(y_j) = \beta_0 + \beta_1 \exp(-\beta_2 x_j)$ “Analysis of running times in Olympic games”
- winning times are decreasing; rate of improvement is decreasing; limiting time (β_0)



```
> plot(year, fifteen, xlab = "year", ylab = "winning time", main = "1500m mens speed
 skating Olympic winning times")
> plot(year, fifteen, xlab = "year", ylab = "winning time", main = "Olympic winning
times", sub = "1500m mens speed skating")
> x <- year-1924
> nls(y ~ b0 + b1*exp(-b2*x), start=(list(b0=110, b1= 1, b2=.5)))
Error in nls(y ~ b0 + b1 * exp(-b2 * x), start = (list(b0 = 110, b1 = 1,  :
  parameters without starting value in 'data': y
> nls(fifteen ~ b0 + b1*exp(-b2*x), start=list(b0=110, b1=1, b2=0.5))
Error in numericDeriv(form[[3L]], names(ind), env) :
  Missing value or an infinity produced when evaluating the model
> nls(fifteen ~ b0 + b1*exp(-b2*x), start=list(b0=100, b1=1, b2=0.5))
Error in numericDeriv(form[[3L]], names(ind), env) :
  Missing value or an infinity produced when evaluating the model
> nls(fifteen ~ b0 + b1*exp(-b2*x), start=list(b0=100, b1=1, b2=0.05))
Error in nls(fifteen ~ b0 + b1 * exp(-b2 * x), start = list(b0 = 100,  :
  singular gradient
> nls(fifteen ~ b0 + b1*exp(-b2*x), start=list(b0=100, b1=1, b2=0.1))
Error in numericDeriv(form[[3L]], names(ind), env) :
  Missing value or an infinity produced when evaluating the model
> nls(fifteen ~ b0 + b1*exp(-b2*x), start=list(b0=100, b1=1, b2=1))
Error in numericDeriv(form[[3L]], names(ind), env) :
  Missing value or an infinity produced when evaluating the model
> nls(fifteen ~ b0 + b1*exp(-b2*x), start=list(b0=100, b1=.1, b2=1))
Error in numericDeriv(form[[3L]], names(ind), env) :
  Missing value or an infinity produced when evaluating the model
> nls(fifteen ~ b0 + b1*exp(-b2*x), start=list(b0=100, b1=.1, b2=.01))
Error in numericDeriv(form[[3L]], names(ind), env) :
  Missing value or an infinity produced when evaluating the model
```

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
-

$$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$

- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
-

$$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$

- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0, 1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
- $$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$
- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
- $$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$
- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \ldots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
- $$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if } \ldots$$
- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \ldots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1)$
  $y_j = 1(z_j > 0)$

Dichotomizing continuous data (§10.4.1)

- suppose $Z_j = x_j^T\gamma + \sigma\epsilon_j, \quad j = 1, \dots, n; \quad \epsilon_j \sim f(\cdot)$
- $Y_j = 1$ if $Z_j > 0$; otherwise 0
- $$\Pr(Y_j = 1) = 1 - F(-x_j^T\gamma/\sigma) = 1 - F(-x_j^T\beta) = F(x_j^T\beta), \text{if ...}$$
- examples (Table 10.7)

logistic	$F(u) = e^u/(1+e^u)$	logit	$log\{p/(1-p)\} = x^T\beta$
normal	$F(u) = \Phi(u)$	probit	$\Phi^{-1}(p) = x^T\beta$
log-Weibull	$F(u) = 1 - \exp(-e^u)$	log-log	$-\log\{-\log(p)\} = x^T\beta$
Gumbel	$F(u) = \exp\{-e^{-u}\}$	c-log-log	$\log\{-\log(1-p)\} = x^T\beta$

- Example 10.17 considers how much information is lost in going from Z to Y
- in special case where $x_j = -1, -0.9, \dots, 0.9, 1$,
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0, 1)$
  $y_j = 1(z_j > 0)$

... example 10.17

- $x_j = -1, -0.9, \dots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = v_Z = 1/\sum(x_i - \bar{x})^2$
- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = v_Y = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = v_Z = 1/\sum(x_i - \bar x)^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = v_Y = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = v_Z = 1/\sum(x_i - \bar x)^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = v_Y = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat{\beta}_Z$ is least squares estimator from original data
- $\text{cov}(\hat{\beta}_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Z}) = v_Z = 1/\sum(x_i - \bar{x})^2$

- $\hat{\beta}_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat{\beta}_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat{\beta}_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Y}) = v_Y = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = v_Z = 1/\sum(x_i - \bar{x})^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = v_Y = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = v_Z = 1/\sum(x_i - \bar x)^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = v_Y = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \dots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat\beta_Z$ is least squares estimator from original data
- $\text{cov}(\hat\beta_Z) = (X^TX)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Z}) = v_Z = 1/\sum(x_i - \bar x)^2$

- $\hat\beta_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat\beta_Y) \doteq (X^TWX)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat\beta_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat\beta_{1Y}) = v_Y = (X^TWX)^{-1}_{(2,2)}$

... example 10.17

- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat{\beta}_Z$ is least squares estimator from original data
- $\text{cov}(\hat{\beta}_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Z}) = v_Z = 1/\sum(x_i - \bar{x})^2$

- $\hat{\beta}_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat{\beta}_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat{\beta}_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Y}) = v_Y = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

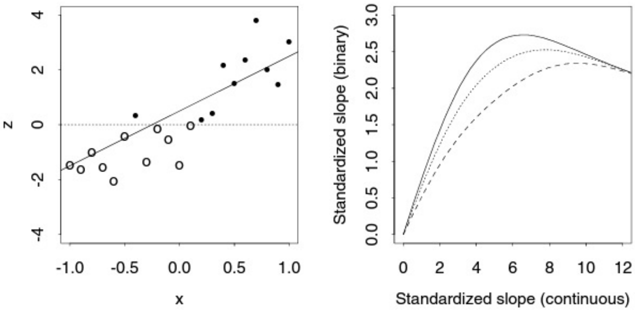
- $x_j = -1, -0.9, \ldots, 0.9, 1,$
  $z_j = 0.5 + 2x_j + \epsilon_j, \quad \epsilon_j \sim N(0,1), \quad y_j = 1(z_j > 0)$
- $\hat{\beta}_Z$ is least squares estimator from original data
- $\text{cov}(\hat{\beta}_Z) = (X^T X)^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Z}) = v_Z = 1/\sum(x_i - \bar{x})^2$

- $\hat{\beta}_Y$ is the estimator from dichotomized data
- $\text{cov}(\hat{\beta}_Y) \doteq (X^T W X)^{-1}, \quad W = \text{diag}(w_j)$ (p.488)
- $w_j = \dfrac{\phi^2(\beta_0 + \beta_1 x_j)}{\Phi(-\beta_0 - \beta_1 x_j)\Phi(\beta_0 + \beta_1 x_j)}$
- $\text{cov}(\hat{\beta}_Y) \doteq \begin{pmatrix} \sum w_j & \sum w_j x_j \\ \sum w_j x_j & \sum w_j x_j^2 \end{pmatrix}^{-1}$
- $\text{var}(\hat{\beta}_{1Y}) = v_Y = (X^T W X)^{-1}_{(2,2)}$

... example 10.17

- Figure 10.6 (right) plots $\beta_1/\sqrt{\sum(x_j - \bar{x})^2} = \beta_1/v_Z$ on the x-axis, and β_1/v_Y on the y-axis
- trying to compare v_Z and v_Y, as well as indicate behaviour of $\beta_{1Y}/\sqrt{v_Y}$ as $\beta_1 \to \infty$

Figure 10.6 Efficiency loss due to reducing continuous variables to binary ones. Left panel: simulated data. Blobs above the dotted line are counted as successes, with zeros below it as failures; the solid line is $0.5 + 2x$. Right panel: Comparison of asymptotic t statistics when continuous data are dichotomized, for normal error distribution, when $\beta_0 = 0.5, 1, 1.5$ (solid, dots, dashes).

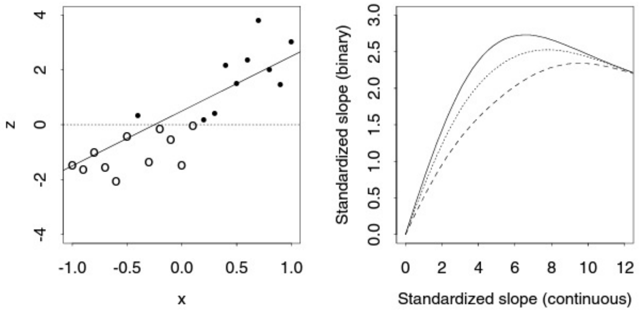


... example 10.17

- Figure 10.6 (right) plots $\beta_1/\sqrt{\sum(x_j - \bar{x})^2} = \beta_1/v_Z$ on the x-axis, and β_1/v_Y on the y-axis
- trying to compare v_Z and v_Y, as well as indicate behaviour of $\beta_{1Y}/\sqrt{v_Y}$ as $\beta_1 \to \infty$

10.4 · Proportion Data 489

Figure 10.6 Efficiency loss due to reducing continuous variables to binary ones. Left panel: simulated data. Blobs above the dotted line are counted as successes, with zeros below it as failures; the solid line is $0.5 + 2x$. Right panel: Comparison of asymptotic t statistics when continuous data are dichotomized, for normal error distribution, when $\beta_0 = 0.5, 1, 1.5$ (solid, dots, dashes).



Binomial regression example

log(dose)	deaths	sample size
-0.86	0	5
-0.30	1	5
-0.05	3	5
0.73	5	5

$$y_j \sim \mathrm{Bin}(5, \pi_j), \quad \log\{\pi_j/(1-\pi_j)\} = \beta_0 + \beta_1 x_j$$

```
> bioassay
      x n r
1 -0.86 5 0
2 -0.30 5 1
3 -0.05 5 3
4  0.73 5 5
> summary(glm(cbind(r,n-r)~x, data = bioassay, family = binomial))

Call:
glm(formula = cbind(r, n - r) ~ x, family = binomial, data = bioassay)

Deviance Residuals:
       1         2         3         4
-0.17236   0.08133  -0.05869   0.12237

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.8466     1.0191   0.831    0.406
x             7.7488     4.8728   1.590    0.112

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 15.791412  on 3  degrees of freedom
Residual deviance:  0.054742  on 2  degrees of freedom
AIC: 7.9648

Number of Fisher Scoring iterations: 7
```

2×2 table

§10.4.2

- special case: $n = 1$ (binary regression)
- covariate takes values 0, 1
- $\Pr(Y_j = 1 \mid x_j = 0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \pi_0$
- $\Pr(Y_j = 1 \mid x_j = 1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} = \pi_1$
- in text: $\psi \leftarrow \beta_1, \lambda \leftarrow \beta_0, T \leftarrow x$
- $Y = 1$ is the event of interest – death, cure, heart attack, ...
- $x = 1$ is the factor of interest – treatment, smoking status, exposure, ... (Davison calls these 'cases')
- it is more usual to call the units with $Y = 1$ the cases (dead, sick, recovered, ...), and $Y = 0$ the controls (alive, well, not recovered ...)

2×2 table

§10.4.2

- special case: $n = 1$ (binary regression)
- covariate takes values 0, 1
- $\Pr(Y_j = 1 \mid x_j = 0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \pi_0$
- $\Pr(Y_j = 1 \mid x_j = 1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} = \pi_1$
- in text: $\psi \leftarrow \beta_1, \lambda \leftarrow \beta_0, T \leftarrow x$
- $Y = 1$ is the event of interest – death, cure, heart attack, ...
- $x = 1$ is the factor of interest – treatment, smoking status, exposure, ... (Davison calls these 'cases')
- it is more usual to call the units with $Y = 1$ the cases (dead, sick, recovered, ...), and $Y = 0$ the controls (alive, well, not recovered ...)

2×2 table

§10.4.2

- special case: $n = 1$ (binary regression)
- covariate takes values 0, 1
- $\Pr(Y_j = 1 \mid x_j = 0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \pi_0$
- $\Pr(Y_j = 1 \mid x_j = 1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} = \pi_1$
- in text: $\psi \leftarrow \beta_1, \lambda \leftarrow \beta_0, T \leftarrow x$
- $Y = 1$ is the event of interest – death, cure, heart attack, ...
- $x = 1$ is the factor of interest – treatment, smoking status, exposure, ... (Davison calls these 'cases')
- it is more usual to call the units with $Y = 1$ the cases (dead, sick, recovered, ...), and $Y = 0$ the controls (alive, well, not recovered ...)

2×2 table

§10.4.2

- special case: $n = 1$ (binary regression)
- covariate takes values 0, 1
- $\Pr(Y_j = 1 \mid x_j = 0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \pi_0$
- $\Pr(Y_j = 1 \mid x_j = 1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} = \pi_1$
- in text: $\psi \leftarrow \beta_1, \lambda \leftarrow \beta_0, T \leftarrow x$
- $Y = 1$ is the event of interest – death, cure, heart attack, ...
- $x = 1$ is the factor of interest – treatment, smoking status, exposure, ... (Davison calls these 'cases')
- it is more usual to call the units with $Y = 1$ the cases (dead, sick, recovered, ...), and $Y = 0$ the controls (alive, well, not recovered ...)

2×2 table

§10.4.2

- special case: $n = 1$ (binary regression)
- covariate takes values 0, 1
- $\Pr(Y_j = 1 \mid x_j = 0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \pi_0$
- $\Pr(Y_j = 1 \mid x_j = 1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} = \pi_1$
- in text: $\psi \leftarrow \beta_1, \lambda \leftarrow \beta_0, T \leftarrow x$

- $Y = 1$ is the event of interest – death, cure, heart attack, ...
- $x = 1$ is the factor of interest – treatment, smoking status, exposure, ... (Davison calls these 'cases')
- it is more usual to call the units with $Y = 1$ the cases (dead, sick, recovered, ...), and $Y = 0$ the controls (alive, well, not recovered ...)

2×2 table

§10.4.2

- special case: $n = 1$ (binary regression)
- covariate takes values 0, 1
- $\Pr(Y_j = 1 \mid x_j = 0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \pi_0$
- $\Pr(Y_j = 1 \mid x_j = 1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} = \pi_1$
- in text: $\psi \leftarrow \beta_1, \lambda \leftarrow \beta_0, T \leftarrow x$

- $Y = 1$ is the event of interest – death, cure, heart attack, ...
- $x = 1$ is the factor of interest – treatment, smoking status, exposure, ... (Davison calls these 'cases')
- it is more usual to call the units with $Y = 1$ the cases (dead, sick, recovered, ...), and $Y = 0$ the controls (alive, well, not recovered ...)

2×2 table

§10.4.2

- special case: $n = 1$ (binary regression)
- covariate takes values 0, 1
- $\Pr(Y_j = 1 \mid x_j = 0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \pi_0$
- $\Pr(Y_j = 1 \mid x_j = 1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} = \pi_1$
- in text: $\psi \leftarrow \beta_1, \lambda \leftarrow \beta_0, T \leftarrow x$

- $Y = 1$ is the event of interest – death, cure, heart attack, ...
- $x = 1$ is the factor of interest – treatment, smoking status, exposure, ... (Davison calls these 'cases')
- it is more usual to call the units with $Y = 1$ the cases (dead, sick, recovered, ...), and $Y = 0$ the controls (alive, well, not recovered ...)

2×2 table

§10.4.2

- special case: $n = 1$ (binary regression)
- covariate takes values 0, 1
- $\Pr(Y_j = 1 \mid x_j = 0) = \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)} = \pi_0$
- $\Pr(Y_j = 1 \mid x_j = 1) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} = \pi_1$
- in text: $\psi \leftarrow \beta_1, \lambda \leftarrow \beta_0, T \leftarrow x$

- $Y = 1$ is the event of interest – death, cure, heart attack, ...
- $x = 1$ is the factor of interest – treatment, smoking status, exposure, ... (Davison calls these 'cases')
- it is more usual to call the units with $Y = 1$ the cases (dead, sick, recovered, ...), and $Y = 0$ the controls (alive, well, not recovered ...)

Prospective and retrospective sampling C &D §3.6

Table 3.6 *Distribution of a binary explanatory variable, z, and response variable, y, in (a) population study, (b) prospective or cohort study, (c) retrospective or case-control study*

(a) Population

	$y=0$	$y=1$
$z=0$	π_{00}	π_{01}
$z=1$	π_{10}	π_{11}

(b) Prospective study

	$y=0$	$y=1$
$z=0$	$\pi_{00}/(\pi_{00}+\pi_{01})$	$\pi_{01}/(\pi_{00}+\pi_{01})$
$z=1$	$\pi_{10}/(\pi_{10}+\pi_{11})$	$\pi_{11}/(\pi_{10}+\pi_{11})$

(c) Retrospective study

	$y=0$	$y=1$
$z=0$	$\pi_{00}/(\pi_{00}+\pi_{10})$	$\pi_{01}/(\pi_{01}+\pi_{11})$
$z=1$	$\pi_{10}/(\pi_{00}+\pi_{10})$	$\pi_{11}/(\pi_{01}+\pi_{11})$

$\pi_{js} = \Pr(z = i, y = s)$, $\quad z$ explanatory, $\quad y$ response

... prospective and retrospective

Population

	$y=0$	$y=1$
$x=0$	π_{00}	π_{01}
$x=1$	π_{10}	π_{11}

Prospective study

	$y=0$	$y=1$
$x=0$	$\pi_{00}/(\pi_{00}+\pi_{01})$	$\pi_{01}/(\pi_{00}+\pi_{01})$
$x=1$	$\pi_{10}/(\pi_{10}+\pi_{11})$	$\pi_{11}/(\pi_{10}+\pi_{11})$

Retrospective study

	$y=0$	$y=1$
$x=0$	$\pi_{00}/(\pi_{00}+\pi_{10})$	$\pi_{01}/(\pi_{01}+\pi_{11})$
$x=1$	$\pi_{10}/(\pi_{00}+\pi_{10})$	$\pi_{11}/(\pi_{01}+\pi_{11})$

odds ratio in 2nd and 3rd table the same be assembled more easily and cheaply than a prospective study, though the lack of randomization weakens subsequent inferences. Let $Z = 1$ indicate that an individual is chosen for the retrospective study, and suppose that this occurs with probabilities

$$\Pr(Z = 1 \mid Y = 1) = p_1, \quad \Pr(Z = 1 \mid Y = 0) = p_0,$$

independent of treatment status T. Then the success probability for an individual who was treated, conditional on their being chosen for inclusion in the study is $\Pr(Y = 1 \mid Z = 1, T = 1)$. This equals

$$\frac{\Pr(Z = 1 \mid Y = 1)\Pr(Y = 1 \mid T = 1)}{\Pr(Z = 1 \mid Y = 1)\Pr(Y = 1 \mid T = 1) + \Pr(Z = 1 \mid Y = 0)\Pr(Y = 0 \mid T = 1)}$$

by Bayes' theorem, so

$$\Pr(Y = 1 \mid Z = 1, T = 1) = \frac{p_1 e^{\lambda+\psi}}{p_1 e^{\lambda+\psi} + p_0} = \frac{e^{\lambda'+\psi}}{1 + e^{\lambda'+\psi}},$$

where $\lambda' = \lambda + \log(p_1/p_0)$. A similar argument gives

$$\Pr(Y = 1 \mid Z = 1, T = 0) = \frac{e^{\lambda'}}{1 + e^{\lambda'}},$$

so although retrospective sampling alters λ, the difference of log odds ψ is unchanged. This gives a strong motivation for using ψ to summarize the treatment effect, particularly if estimates from both types of study will ultimately be combined.

This argument applies also if ψ is replaced by $x^{\mathrm{T}}\beta$, where x contains covariates as well as an indicator of treatment status. The key point is that the selection probabilities p_1 and p_0 must be independent of x.

Contingency Tables

Example 10.19

Age (years)	Smokers	Non-smokers
Overall	139/582 (24)	230/732 (31)
18–24	2/55 (4)	1/62 (2)
25–34	3/124 (2)	5/157 (3)
35–44	14/109 (13)	7/121 (6)
45–54	27/130 (21)	12/78 (15)
55–64	51/115 (44)	40/121 (33)
65–74	29/36 (81)	101/129 (78)
75+	13/13 (100)	64/64 (100)

Table 6.8 Twenty-year survival and smoking status for 1314 women (Appleton *et al.*, 1996). The smoker and non-smoker columns contain number dead/total (% dead).

	Smoker	Non-smoker	
dead	139 (24%)	230 (31%)	
alive	443	502	
total	582	732	1314

... Example 10.19

```
>  summary(glm(cbind(alive,dead) ~ smoker, data = smoking, family = binomial))
Call:
glm(formula = cbind(alive, dead) ~ smoker, family = binomial,
    data = smoking)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-12.173   -5.776    1.869    5.674    9.052

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.78052    0.07962   9.803  < 2e-16 ***
smoker       0.37858    0.12566   3.013  0.00259 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 641.5  on 13  degrees of freedom
Residual deviance: 632.3  on 12  degrees of freedom
AIC: 683.29

Number of Fisher Scoring iterations: 4
```

... example 10.19

	Smoker	Non-smoker	
dead	139 (24%)	230 (31%)	
alive	443	502	
total	582	732	1314

```
 > anova(glm(cbind(alive,dead) ~ smoker, data = smoking, family = binomial))
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(alive, dead)

Terms added sequentially (first to last)


       Df Deviance Resid. Df Resid. Dev
NULL                      13      641.5
smoker  1   9.2003        12      632.3
> with(smoking, xtabs(cbind(dead,alive) ~ smoker))

smoker dead alive
     0  230   502
     1  139   443
> summary(.Last.value)
Call: xtabs(formula = cbind(dead, alive) ~ smoker)
Number of cases in table: 1314
Number of factors: 2
Test for independence of all factors:
Chisq = 9.121, df = 1, p-value = 0.002527
```

... Example 10.19

	sm	non-sm	sm	non-sm	sm	non-sm	
d	2	1	3	5	14	7	
a	53	61	121	152	95	114	...
	55	62	124	157	109	121	
Age	18-24		25-34		35-44		...

```
> summary(glm(cbind(alive,dead) ~ smoker + factor(age), data = smoking, family = binomial))
...
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)         3.8601     0.5939   6.500 8.05e-11 ***
smoker             -0.4274     0.1770  -2.414 0.015762 *
factor(age)25-34   -0.1201     0.6865  -0.175 0.861178
factor(age)35-44   -1.3411     0.6286  -2.134 0.032874 *
factor(age)45-54   -2.1134     0.6121  -3.453 0.000555 ***
factor(age)55-64   -3.1808     0.6006  -5.296 1.18e-07 ***
factor(age)65-74   -5.0880     0.6195  -8.213  < 2e-16 ***
factor(age)75+    -27.8073 11293.1437  -0.002 0.998035
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 641.4963  on 13  degrees of freedom
Residual deviance:   2.3809  on  6  degrees of freedom
AIC: 65.377

Number of Fisher Scoring iterations: 20
```

Log-linear models 10.5.2

- suppose we have 3 factors, each with several levels
- observe a response at each combination of factors
- linear model might be

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \quad k = 1, \dots, K; j = 1, \dots, J; i = 1, \dots I$$

- or

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

- if the y_{ijk} are positive counts, rather than continuous, then Poisson model could have

$$y_{ijk} \sim Po(\mu_{ijk}), \quad \log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k$$

- or

$$\log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

Log-linear models 10.5.2

- suppose we have 3 factors, each with several levels
- observe a response at each combination of factors
- linear model might be

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \quad k = 1, \dots, K; j = 1, \dots, J; i = 1, \dots I$$

- or

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

- if the y_{ijk} are positive counts, rather than continuous, then Poisson model could have

$$y_{ijk} \sim Po(\mu_{ijk}), \quad \log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k$$

- or

$$\log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

Log-linear models

10.5.2

- suppose we have 3 factors, each with several levels
- observe a response at each combination of factors
- linear model might be

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \quad k = 1, \ldots, K; j = 1, \ldots, J; i = 1, \ldots I$$

- or

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

- if the y_{ijk} are positive counts, rather than continuous, then Poisson model could have

$$y_{ijk} \sim Po(\mu_{ijk}), \quad \log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k$$

- or

$$\log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

Log-linear models

10.5.2

- suppose we have 3 factors, each with several levels
- observe a response at each combination of factors
- linear model might be

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \quad k = 1, \ldots, K; j = 1, \ldots, J; i = 1, \ldots I$$

- or

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

- if the y_{ijk} are positive counts, rather than continuous, then Poisson model could have

$$y_{ijk} \sim Po(\mu_{ijk}), \quad \log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k$$

- or

$$\log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

Log-linear models

- suppose we have 3 factors, each with several levels
- observe a response at each combination of factors
- linear model might be

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \quad k = 1, \dots, K; j = 1, \dots, J; i = 1, \dots I$$

- or

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

- if the y_{ijk} are positive counts, rather than continuous, then Poisson model could have

$$y_{ijk} \sim Po(\mu_{ijk}), \quad \log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k$$

- or

$$\log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

Log-linear models

- suppose we have 3 factors, each with several levels
- observe a response at each combination of factors
- linear model might be

$$y_{ijk} = \mu+\alpha_i+\beta_j+\gamma_k+\epsilon_{ijk}, \quad k = 1, \dots, K; j = 1, \dots, J; i = 1, \dots I$$

- or

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

- if the y_{ijk} are positive counts, rather than continuous, then Poisson model could have

$$y_{ijk} \sim Po(\mu_{ijk}), \quad \log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k$$

- or

$$\log(\mu_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

Example 10.23

8 1 · *Introduction*

Years of smoking t	Daily cigarette consumption d						
	Nonsmokers	1–9	10–14	15–19	20–24	25–34	35+
15–19	10366/1	3121	3577	4317	5683	3042	670
20–24	8162	2937	3286/1	4214	6385/1	4050/1	1166
25–29	5969	2288	2546/1	3185	5483/1	4290/4	1482
30–34	4496	2015	2219/2	2560/4	4687/6	4268/9	1580/4
35–39	3512	1648/1	1826	1893	3646/5	3529/9	1336/6
40–44	2201	1310/2	1386/1	1334/2	2411/12	2424/11	924/10
45–49	1421	927	988/2	849/2	1567/9	1409/10	556/7
50–54	1121	710/3	684/4	470/2	857/7	663/5	255/4
55–59	826/2	606	449/3	280/5	416/7	284/3	104/1

Table 1.4 Lung cancer deaths in British male physicians (Frome, 1983). The table gives man-years at risk/number of cases of lung cancer, cross-classified by years of smoking, taken to be age minus 20 years, and number of cigarettes smoked per day.

```
> data(lung.cancer)
> lung.cancer[1:3,]
  years.smok cigarettes  Time y
1      15-19          0 10366 1
2      15-19        1-9  3121 0
3      15-19      10-14  3577 0
```

... example 10.23

Years of smoking t	Daily cigarette consumption d						
	Nonsmokers	1–9	10–14	15–19	20–24	25–34	35+
15–19	10366/1	3121	3577	4317	5683	3042	670
20–24	8162	2937	3286/1	4214	6385/1	4050/1	1166
25–29	5969	2288	2546/1	3185	5483/1	4290/4	1482
30–34	4496	2015	2219/2	2560/4	4687/6	4268/9	1580/4
35–39	3512	1648/1	1826	1893	3646/5	3529/9	1336/6
40–44	2201	1310/2	1386/1	1334/2	2411/12	2424/11	924/10
45–49	1421	927	988/2	849/2	1567/9	1409/10	556/7
50–54	1121	710/3	684/4	470/2	857/7	663/5	255/4
55–59	826/2	606	449/3	280/5	416/7	284/3	104/1

Table 1.4 Lung cancer deaths in British male physicians (Frome, 1983). The table gives man-years at risk/number of cases of lung cancer, cross-classified by years of smoking, taken to be age minus 20 years, and number of cigarettes smoked per day.

$$y_{td} \sim \text{Poisson}(T_{td}\mu_{td}) \qquad T_d = \text{man-years}$$

$$\mu_{td} = \exp(\alpha_t + \beta_d)$$

... example 10.23

Years of smoking t	Daily cigarette consumption d						
	Nonsmokers	1–9	10–14	15–19	20–24	25–34	35+
15–19	10366/1	3121	3577	4317	5683	3042	670
20–24	8162	2937	3286/1	4214	6385/1	4050/1	1166
25–29	5969	2288	2546/1	3185	5483/1	4290/4	1482
30–34	4496	2015	2219/2	2560/4	4687/6	4268/9	1580/4
35–39	3512	1648/1	1826	1893	3646/5	3529/9	1336/6
40–44	2201	1310/2	1386/1	1334/2	2411/12	2424/11	924/10
45–49	1421	927	988/2	849/2	1567/9	1409/10	556/7
50–54	1121	710/3	684/4	470/2	857/7	663/5	255/4
55–59	826/2	606	449/3	280/5	416/7	284/3	104/1

Table 1.4 Lung cancer deaths in British male physicians (Frome, 1983). The table gives man-years at risk/number of cases of lung cancer, cross-classified by years of smoking, taken to be age minus 20 years, and number of cigarettes smoked per day.

$$y_{td} \sim \text{Poisson}(T_{td}\mu_{td}) \qquad T_d = \text{man-years}$$

$$\mu_{td} = \exp(\alpha_t + \beta_d)$$

… example 10.23

8   *1 · Introduction*

Years of smoking t	Daily cigarette consumption d						
	Nonsmokers	1–9	10–14	15–19	20–24	25–34	35+
15–19	10366/1	3121	3577	4317	5683	3042	670
20–24	8162	2937	3286/1	4214	6385/1	4050/1	1166
25–29	5969	2288	2546/1	3185	5483/1	4290/4	1482
30–34	4496	2015	2219/2	2560/4	4687/6	4268/9	1580/4
35–39	3512	1648/1	1826	1893	3646/5	3529/9	1336/6
40–44	2201	1310/2	1386/1	1334/2	2411/12	2424/11	924/10
45–49	1421	927	988/2	849/2	1567/9	1409/10	556/7
50–54	1121	710/3	684/4	470/2	857/7	663/5	255/4
55–59	826/2	606	449/3	280/5	416/7	284/3	104/1

Table 1.4 Lung cancer deaths in British male physicians (Frome, 1983). The table gives man-years at risk/number of cases of lung cancer, cross-classified by years of smoking, taken to be age minus 20 years, and number of cigarettes smoked per day.

$$y_{td} \sim \text{Poisson}(T_{td}\mu_{td}) \qquad T_d = \text{man-years}$$

$$\mu_{td} = \exp(\alpha_t + \beta_d)$$

... Example 10.23

```
> summary(glm(y ~ cigarettes + years.smok + offset(log(Time)),
family = poisson, data = lung.cancer))
...

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      -12.5784     1.1475 -10.961  < 2e-16 ***
cigarettes1-9      1.2200     0.7073   1.725 0.084547 .
cigarettes10-14    2.0991     0.6363   3.299 0.000971 ***
cigarettes15-19    2.3089     0.6327   3.649 0.000263 ***
cigarettes20-24    2.9009     0.5956   4.870 1.11e-06 ***
cigarettes25-34    3.1162     0.5947   5.240 1.61e-07 ***
cigarettes35+      3.6059     0.6048   5.962 2.49e-09 ***
...

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 445.099  on 62  degrees of freedom
Residual deviance:  51.471  on 48  degrees of freedom
AIC: 201.31
```

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d, t) = (\beta_0 + \beta_1 d^{\beta_2}) t^{\beta_3}$
- if $\beta_2 = 0$, β_1 not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d, t) = \{e^{\gamma_0} \exp(\gamma_1 + \beta_2 \log d)\} \exp(\beta_3 \log t)$
- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (y is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with R; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d,t) = (\beta_0 + \beta_1 d^{\beta_2}) t^{\beta_3}$
- if $\beta_2 = 0$, β_1 not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d,t) = \{e^{\gamma_0} \exp(\gamma_1 + \beta_2 \log d)\} \exp(\beta_3 \log t)$
- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (y is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with R; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d, t) = (\beta_0 + \beta_1 d^{\beta_2}) t^{\beta_3}$
- if $\beta_2 = 0, \beta_1$ not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d, t) = \{e^{\gamma_0} \exp(\gamma_1 + \beta_2 \log d)\} \exp(\beta_3 \log t)$
- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (y is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with R; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d, t) = (\beta_0 + \beta_1 d^{\beta_2}) t^{\beta_3}$
- if $\beta_2 = 0, \beta_1$ not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d, t) = \{e^{\gamma_0} \exp(\gamma_1 + \beta_2 \log d)\} \exp(\beta_3 \log t)$
- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (y is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with R; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d,t) = (\beta_0 + \beta_1 d^{\beta_2})t^{\beta_3}$

- if $\beta_2 = 0, \beta_1$ not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d,t) = \{e^{\gamma_0}\exp(\gamma_1 + \beta_2 \log d)\}\exp(\beta_3 \log t)$

- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (y is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with R; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d,t) = (\beta_0 + \beta_1 d^{\beta_2}) t^{\beta_3}$

- if $\beta_2 = 0, \beta_1$ not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d,t) = \{e^{\gamma_0} \exp(\gamma_1 + \beta_2 \log d)\} \exp(\beta_3 \log t)$

- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (y is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with R; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d, t) = (\beta_0 + \beta_1 d^{\beta_2})t^{\beta_3}$
- if $\beta_2 = 0, \beta_1$ not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d, t) = \{e^{\gamma_0} \exp(\gamma_1 + \beta_2 \log d)\} \exp(\beta_3 \log t)$
- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (y is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with R; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d,t) = (\beta_0 + \beta_1 d^{\beta_2}) t^{\beta_3}$

- if $\beta_2 = 0, \beta_1$ not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d,t) = \{e^{\gamma_0} \exp(\gamma_1 + \beta_2 \log d)\} \exp(\beta_3 \log t)$

- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (*y* is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with `R`; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)

... log-linear models

- p.501, fit a 4-parameter model
- $\lambda(d, t) = (\beta_0 + \beta_1 d^{\beta_2}) t^{\beta_3}$

- if $\beta_2 = 0, \beta_1$ not estimable; similarly if $\beta_1 = 0$
- reparameterize to $\lambda(d, t) = \{e^{\gamma_0} \exp(\gamma_1 + \beta_2 \log d)\} \exp(\beta_3 \log t)$

- model fits quite well, fewer estimated parameters, $\beta_2 = 1$ corresponds to linear growth
- see also Example 10.21 for a Poisson example (y is number of goals scored in soccer match)
- with the Poisson-multinomial connection, we can also fit contingency tables with more than one response factor
- Faraway: Extending the Linear Model with R; Agresti: Analysis of Categorical Data
- skip "marginal models" (p.505) and ordinal data (§10.5.2)