Today

- HW 2 due March 4
- Case Studies,SSC Annual Meeting
- model choice
- random effects and mixed effects models ELM Ch. 8
- generalized linear models separate systematic part of the model from the random part of the model

- linear predictor: $g(\mu_i) = x_i^{\mathrm{T}}\beta$ $\quad \mathrm{E}(y_i) = \mu_i; \quad \mathrm{var}(y_i) = \phi V(\mu_i)$
- exponential family: $f(y_i; \mu_i) \propto \exp[\{\theta_i y_i - b(\theta_i)\}/(a_i\phi) + c(y_i, \phi)]$

- model choice concerns how to build the linear predictor

  linear in β
- nonlinear least squares generalizes η, keeps $f(\cdot)$ in a small class location: normal, sometimes t, occasionally extreme-value

Empirical models

CD Ch. 6.5

- in many fields of study the models used as a basis for interpretation do not have a special subject-matter base
- rather represent broad patterns of haphazard variation quite widely seen
- this is typically combined with a specification of the systematic part of the variation
- which is often the primary focus
- modelling then often reduces to a choice of distributional form
- and of the independence structure of the random components

... empirical models

CD Ch. 6.5

- functional form of the probability distribution sometimes critical, for example where an implicit assumption is involved of a relationship between variance and mean: geometric, Poisson, binomial
- the simple situations that give rise to binomial, Poisson, geometric, exponential, normal and log normal are some guide to empirical model choice in more complex situations
- In some specific contexts there is a tradition establishing the form of model likely to be suitable
- illustration: financial time series – $Y(t) = \log\{P(t)/P(t-1)\}$ has a long-tailed distribution, small serial correlation, large serial correlation in $Y^2(t)$
- illustration: a common type of response arises as the time from some clearly defined origin to a critical event
- often have a long tail of large values; exponential distribution is a natural staring point
- extensions may be needed, including Weibull, gamma or log-normal

... empirical models

CD Ch. 6.5

- often helpful to develop random and systematic parts of the model separately
- models should obey natural or known constraints, even if these lie outside the range of the data
- example $P(Y = 1) = \alpha + \beta x$
- often use instead $\log \frac{P(Y=1)}{P(Y=0)} = \alpha' + \beta' x$
- however, β measures the change in probability per unit change in x
- in many common applications, relationship between y and several variables $x_1, \ldots x_p$ is involved
  - unlikely that the system is wholly linear
  - impractical to study nonlinear systems of unknown form
  - therefore reasonable to begin with a linear model
  - and seek isolated nonlinearities

... empirical models

CD Ch. 6.5

- often helpful to develop random and systematic parts of the model separately
- naive approach: one random variable per study individual
- values for different individuals independent
- more realistic: possibility of structure in the random variation
- dependence in time or space, or a hierarchical structure corresponding to levels of aggregation
- ignoring these complications may give misleading assessments of precision, or bias the conclusions

... empirical models

CD Ch. 6.5

- example: standard error of mean $\sigma/\sqrt{n}$
- but, under mutual correlation, becomes $(\sigma/\sqrt{n})(1+\Sigma\rho_{ij})^{1/2}$
- if each observation correlated with k others, at same level, $(\sigma/\sqrt{n})(1+k\rho)^{1/2}$
- ```
  0.1  0.2  0.4  0.8
  --------------------------
  1.14 1.26 1.48 1.84
  1.18 1.34 1.61 2.05
  1.22 1.41 1.73 2.24
  1.26 1.48 1.84 2.41
  1.30 1.55 1.95 2.57
  1.34 1.61 2.05 2.72
  ```
  ```

... empirical models

CD Ch. 6.5

- important to be explicit about the unit of analysis
- has a bearing on independence assumptions involved in model formulation
- example: if all patients in the same clinic receive the same treatment
- then the clinic is the unit of analysis
- in some contexts there may be a clear hierarchy
- assessment of precision comes primarily from comparisons between units of analysis
- modelling of variation within units is necessary only if of intrinsic interest
- when relatively complex responses are collected on each study individual, the simplest way of condensing these is through a number of summary descriptive measures
- in other situations it may be necessary to represent explicitly the different hierarchies of variation

Models with random effects

- simplest case: one-way layout, linear model

  comparing *a* groups; equality of means

- $y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \dots, a; \quad j = 1, \dots n$
- usually assume $\epsilon_{ij} \sim N(0, \sigma^2)$
- ANOVA:

Source	df	SS	MS	E(MS)
between groups	$a-1$	$\Sigma_{ij}(\bar{y}_{i.} - \bar{y}_{..})^2$	SS_b/df_b	$\sigma^2 + \frac{n\Sigma_i\alpha_i^2}{a-1}$
within groups	$a(n-1)$	$\Sigma_{ij}(y_{ij} - \bar{y}_{i.})^2$	SS_w/df_w	σ^2

- MS_b/MS_w follows an $F_{(a-1),a(n-1)}$ distribution under $H_0 : \alpha_i = 0, i = 1, \dots, a$

... random effects

- change the model assumptions
- $y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \dots, a; \quad j = 1, \dots n$
- $\alpha_i \sim N(0, \sigma_a^2), \quad \epsilon_{ij} \sim N(0, \sigma^2)$
- ANOVA:

Source	df	SS	MS	E(MS)
between groups	$a-1$	$\Sigma_{ij}(\bar{y}_{i.} - \bar{y}_{..})^2$	SS_b/df_b	$\sigma^2 + n\sigma_a^2$
within groups	$a(n-1)$	$\Sigma_{ij}(y_{ij} - \bar{y}_{i.})^2$	SS_w/df_w	σ^2

- MS_b/MS_w follows an $F_{(a-1),a(n-1)}$ distribution under $H_0 : \sigma_a^2 = 0$

Inference

- fixed effects model
- $\text{var}(\bar{y}_{i.} - \bar{y}_{i'.}) = 2\sigma^2/n$
- confidence intervals for $\mu_i - \mu_{i'}$
- σ^2 needs to be estimated, but not of particular interest
- typically use $MSE = SSE/\{a(n-1)\}$

- random effects model
- The parameters σ^2 and σ_a^2 are now of interest
- $\tilde{\sigma}^2 = MSE; \quad \tilde{\sigma}_a^2 = ?$

- maximum likelihood estimates
- REML: restricted maximum likelihood estimates

Another easy example: two-way layout

- randomized block design
- $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, \dots, a; j = 1, \dots, b$
- $\beta_j \sim N(0, \sigma_b^2), \epsilon_{ij} \sim N(0, \sigma^2)$
- a mixed effect model, with one fixed effect (treatment) and one random effect (blocks)
- ANOVA:

Source	df	SS	E(MS)
treatments	$a-1$	$\Sigma_{ij}(\bar{y}_{i.} - \bar{y}_{..})^2$	$\sigma^2 + \frac{n\Sigma_i\alpha_i^2}{a-1}$
blocks	$b-1$	$\Sigma_{ij}(\bar{y}_{.j} - \bar{y}_{..})^2$	$\sigma^2 + a\sigma_b^2$
error	$(a-1)(b-1)$	$\Sigma_{ij}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	σ^2

$$\mathrm{cov}(y_{ij}, y_{i'j}) = \mathrm{cov}(\beta_j + \epsilon_{ij}, \beta_j + \epsilon_{i'j}) = \sigma_b^2 + \sigma^2$$

Randomized block design with repeats

- repeat observations for each treatment, in each block
- $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$
  $i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots n$
- $\beta_j \sim N(0, \sigma_b^2), (\alpha\beta)_{ij} \sim N(0, \sigma_{ab}^2), \epsilon_{ij} \sim N(0, \sigma^2)$
- ANOVA:

Source	df	SS	E(MS)
treatments	$a-1$	$\Sigma_{ijk}(\bar{y}_{i.} - \bar{y}_{..})^2$	$\sigma^2 + n\sigma_{ab}^2 + \frac{nb\Sigma_i\alpha_i^2}{a-1}$
blocks	$b-1$	$\Sigma_{ijk}(\bar{y}_{.j} - \bar{y}_{..})^2$	$\sigma^2 + na\sigma_b^2$
interaction	$(a-1)(b-1)$	$\Sigma_{ijk}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$\sigma^2 + n\sigma_{ab}^2$
error	$(n-1)ab$	$\Sigma_{ijk}(y_{ijk} - \bar{y}_{ij.})^2$	σ^2

if the repeats are 'true replications', then we have a full factorial

A general framework

$$y \mid \gamma = X\beta + Z\gamma + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$$

- γ a $q-$vector of random effects; β a p-vector of fixed effects
- assumption $\gamma \sim N(0, \sigma^2 D)$

- marginal distribution

$$y \sim N(X\beta, \sigma^2(I + ZDZ^{\mathrm{T}})) = N(X\beta, \sigma^2 V), \text{say}$$

- applications
  - multi-level models
  - repeated measures
  - longitudinal data
  - components of variance

SM Example 9.16

Example 9.16 (Longitudinal data) A short longitudinal study has one individual allocated to the treatment and two to the control, with observations

$$y_{1j} = \beta_0 + b_1 + \varepsilon_{1j}, \quad y_{21} = \beta_0 + b_2 + \varepsilon_{21}, \quad y_{3j} = \beta_0 + \beta_1 + b_3 + \varepsilon_{3j}, \quad j = 1, 2.$$

Thus there are two measurements on the first and third individuals, and just one on the second. The b_j represent variation among individuals and the ε_{ij} variation between measures on the same individuals. If the b's and ε's are all mutually independent with variances σ_b^2 and σ^2, then

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{pmatrix},$$

and this fits into formulation (9.12) with $\Omega_b = \sigma_b^2 I_3$ and $\Omega = \sigma^2 I_5$. Here ψ comprises the scalar σ_b^2/σ^2, and hence the variance matrix

$$\Omega + Z\Omega_b Z^{\mathrm{T}} = \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_b^2 + \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_b^2 + \sigma^2 & \sigma_b^2 \\ 0 & 0 & 0 & \sigma_b^2 & \sigma_b^2 + \sigma^2 \end{pmatrix}$$

may be written as

$$\begin{pmatrix} 1+\psi & \psi & 0 & 0 & 0 \end{pmatrix}$$

Estimation

- $y \sim N(X\beta, \sigma^2(I + ZDZ^{\mathrm{T}})) = N(X\beta, \sigma^2 V)$
- $$\ell(\beta; y) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log|V| - \frac{1}{2\sigma^2}(y - X\beta)^{\mathrm{T}} V^{-1}(y - X\beta)$$
- V may have one or more unknown parameters
- Example 9.16: $\gamma \sim N_3(0, \sigma_b^2 I)$
- $$I{+}ZDZ^{\mathrm{T}} = \begin{pmatrix} 1+\sigma_b^2/\sigma^2 & \sigma_b^2/\sigma^2 & 0 & 0 & 0 \\ \sigma_b^2/\sigma^2 & 1+\sigma_b^2/\sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 1+\sigma_b^2/\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 1+\sigma_b^2/\sigma^2 & \sigma_b^2/\sigma^2 \\ 0 & 0 & 0 & \sigma_b^2/\sigma^2 & 1+\sigma_b^2/\sigma^2 \end{pmatrix}$$
- $\hat{\beta}_\psi = (X^{\mathrm{T}} V^{-1} X)^{-1} X^{\mathrm{T}} V^{-1} y$
  $\hat{\sigma}_\psi^2 = \frac{1}{n}(y - X\hat{\beta}_\psi)^{\mathrm{T}} V^{-1}(y - X\hat{\beta}_\psi)$

... estimation

- $\hat{\beta}_\psi = (X^{\mathrm{T}} V^{-1} X)^{-1} X^{\mathrm{T}} V^{-1} y$
  $\hat{\sigma}^2_\psi = \frac{1}{n}(y - X\hat{\beta}_\psi)^{\mathrm{T}} V^{-1}(y - X\hat{\beta}_\psi)$

- profile log-likelihood

$$\ell_{\mathrm{p}}(\psi) = -\frac{1}{2}\log \hat{\sigma}^2_\psi - \frac{1}{2}\log |V_\psi|$$

- to get better divisors — properly adjust for degrees of freedom
- modified profile log-likelihood

also called restricted profile log-likelihood

$$\begin{aligned} \ell_{\mathrm{mp}}(\sigma^2, \psi) &= -\frac{1}{2}\log |V_\psi| - \frac{1}{2}\log |X^{\mathrm{T}} V_\psi^{-1} X| \\ &\quad -\frac{1}{2\sigma^2}(y - X\hat{\beta}_\psi)^{\mathrm{T}} V_\psi^{-1}(y - X\hat{\beta}_\psi) - \frac{n-p}{2}\log \sigma^2 \end{aligned}$$

- $$\ell_{\mathrm{p}}(\sigma^2, \psi) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log |V| - \frac{1}{2\sigma^2}\hat{\sigma}^2_\psi$$

Example 9.18

- repeated measurements on the 30 individuals, at 5 time points
- might expect that regression relationship against time is similar for each individual, subject to random variation
- model $y_{jt} = \beta_0 + b_{j0} + (\beta_1 + b_{j1})x_{jt} + \epsilon_{jt}, \quad t = 1, \dots, 5$
- x_{jt} takes values $0, 1, 2, 3, 4$ for $t = 1, 2, 3, 4, 5$
- same for each j
- `data(rat.growth, library="SMPracticals")`
- $(b_{j0}, b_{j1}) \dot\sim N_2(0, \Omega_b), \quad \epsilon_{jt} \dot\sim N(0, \sigma^2)$ independent
- two fixed parameters β_0, β_1
- four variance/covariance parameters: $\sigma^2_{b0}, \sigma^2_{b1}, \text{cov}(b_0, b_1), \sigma^2$

... Example 9.18

- maximum likelihood estimates of fixed effects: $\hat{\beta}_0 = 156.05(2.16), \hat{\beta}_1 = 43.27(0.73)$
- weight in week 1 is estimated to be about 156 units, and average increase per week estimated to be 43.27
- there is large variability between rats: estimated standard deviation of 10.93 for intercept, 3.53 for slope
- there is little correlation between the intercepts and slopes
- ```
  library(MASS)  # this is included the standard R distribution
  library(SMPracticals) # this has various data sets from Davison's book
  library(ellipse) # but I got an error the first time and had to download an additional p
  library(SMPracticals) # and now it works
  data(rat.growth) # for Example 9.18
  rat.growth[1:10,]  # to see what it looks like, and to see variable names
  with(rat.growth, plot( y ~ week , type="l"))
  separate.lm = lm(y ~ week + factor(rat)+ week:factor(rat), data = rat.growth) # fit sep
  rat.mixed = lmer(y ~ week + (week|rat), data = rat.growth) # REML is the default
  summary(rat.mixed)  # compare Table 9.28
  ```
  ```