Multilevel Linear Models: Varying Slopes, Non-Nested Models, and Other Complexities

JEFF GILL
Distinguished Professor, Department of Government
Professor, Department of Mathematics & Statistics
Member, Center for Neuroscience
American University

Credible Intervals and Sets

- The Bayesian analogue to the confidence interval is the credible interval and more generally the credible set, which does not have to be contiguous.
- Most of the time in practice, it is calculated in *exactly the same way* as the confidence interval.
- For instance calculating a 95% credible interval under the Gaussian normal assumption means marching-out 1.96 standard errors from the mean in either direction, just like the analagous confidence interval is created. (The difference lies in the interpretation.)
- A $100(1-\alpha)\%$ credible interval gives the region of the parameter space where the probability of covering $\boldsymbol{\theta}$ is at least $1-\alpha$.
- In contrast, applying this new definition to the confidence interval means that the probability of coverage is either zero or one, since it either covers the true $\boldsymbol{\theta}$ or it doesn't.

Credible Intervals and Sets

- Define C as a *contiguous* subset of the parameter space, $\boldsymbol{\Theta}$, such that a $100(1-\alpha)$ credible interval meets the condition:

$$1-\alpha = \int_C \pi(\boldsymbol{\theta}|\mathbf{X})d\theta$$

  for some chosen α level.

- Conventions: centered at mean or mode, equal tails.

- So credible intervals are *not* necessarily unique!

Credible Intervals and Sets, Example

- Suppose we have duration data, $\mathbf{X}$, exponentially distributed $p(X|\theta) = \theta e^{-\theta X}$ defined over $(0, \infty)$, where interest is in the posterior distribution of the unknown parameter θ.
- Specify the prior distribution of $p(\theta) = 1/\theta$, for $\theta \in (0{:}\infty)$.
- The posterior is:

$$\pi(\theta|\mathbf{X}) \propto p(\theta)L(\theta|\mathbf{X}) = \frac{1}{\theta}\theta^n \exp\left[-\theta \sum_{i=1}^{n} x_i\right] = \theta^{n-1} \exp\left[-\theta \sum_{i=1}^{n} x_i\right].$$

- This means that $\theta|\mathbf{X} \sim \mathcal{G}(\theta|n, \sum x_i)$, where putting the constants back in front to recover the full form of this gamma posterior distribution produces:

$$\pi(\theta|\mathbf{X}) = \frac{(\sum x_i)^n}{\Gamma(n)}\theta^{n-1} \exp\left[-\theta \sum x_i\right].$$

- Since we know everything about this posterior distribution, we are free to choose any desired credible interval.

Credible Intervals and Sets, Example

- Browne, Freidreis, and Gleiber (1986) tabulate complete cabinet duration for eleven Western European countries from 1945 to 1980:

Table 1: European Cabinet Duration Annualized, 1945-1980

Country	N	Average Duration
Austria	15	2.114
Belgium	27	1.234
Denmark	20	1.671
Finland	28	1.070
Iceland	15	2.080
Ireleand	14	2.629
Italy	38	0.833
Netherland	12	2.637
Norway	17	2.065
Sweden	15	2.274

Credible Intervals and Sets, Example

- Country averages from the third column of the table are weighted by N in the second column to reflect the number of such events: $\mathbf{X}_i N_i$.
- For a chosen α the end-points of an equal-tail credible interval can be calculated with:

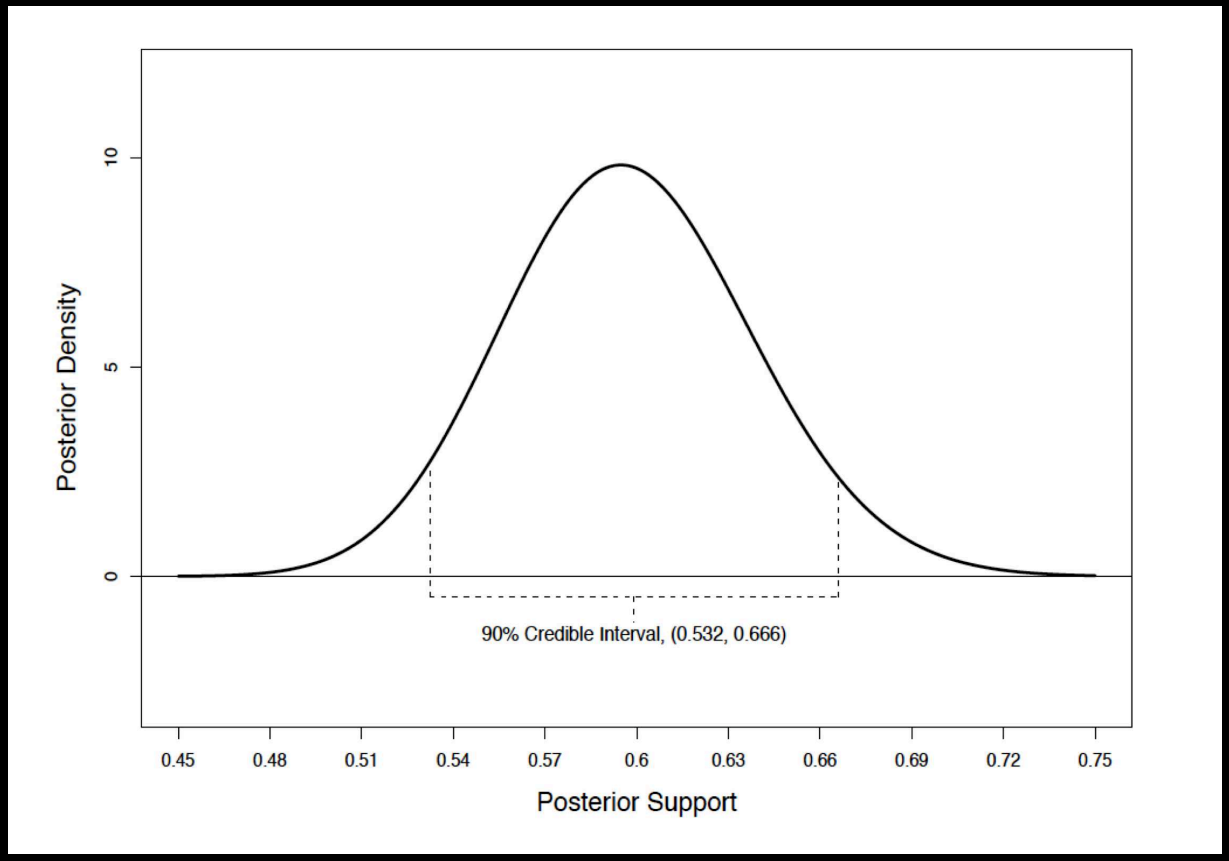
$$\frac{\alpha}{2} = \int_0^L \pi(\boldsymbol{\theta}|\mathbf{X})d\theta \qquad \frac{\alpha}{2} = \int_H^\infty \pi(\boldsymbol{\theta}|\mathbf{X})d\theta$$

or we could simply use the following R commands for a 95% credible interval:

```
dur <- c(2.114,1.234,1.671,1.070,2.168,2.080, 2.629,0.833,2.637,2.065,2.274)
N <- c(15,27,20,28,15,15,14,38,12,17,15)
qgamma(0.025,shape=sum(N),rate=sum(N*dur))
[1] 0.52056
qgamma(0.975,shape=sum(N),rate=sum(N*dur))
[1] 0.67988
```

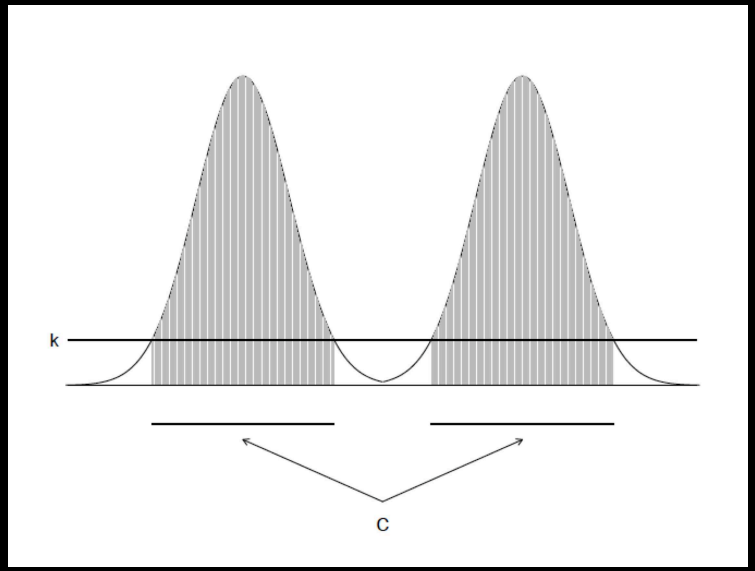
Credible Intervals and Sets, Example

Equal Tail Credible Interval for Cabinet Duration



Highest Posterior Density Intervals and Sets

- When looking at posterior distributions, we really care where the highest density exists on the support of the posterior density, regardless of whether it is contiguous or not.
- HPD created such that that no region outside of the interval will have higher posterior density than any region inside the the HPD.
- Therefore HPDs are not necessarily contiguous.



Highest Posterior Density Intervals and Sets

- A $100(1-\alpha)\%$ highest posterior density (HPD) is the subset of the support of the posterior distribution for some parameter, θ, that meets the criteria:

$$C = \{\theta : \pi(\theta|\mathbf{x}) \geq k\},$$

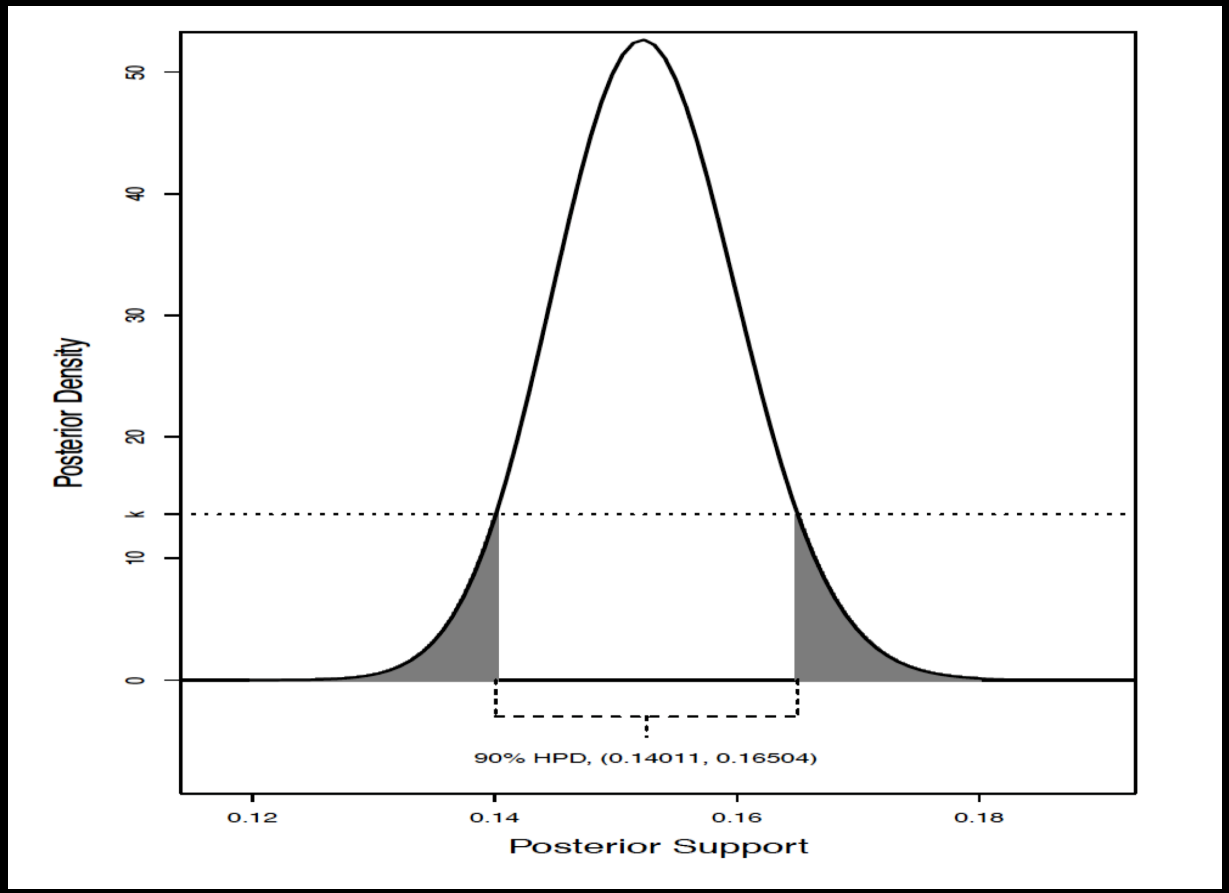
where k is the largest number such that:

$$1-\alpha = \int_{\theta:\pi(\theta|\mathbf{x})\geq k} \pi(\theta|\mathbf{x})d\theta$$

- The important difference is $\theta : \pi(\theta|\mathbf{x}) \geq k$ instead of a single contiguous interval as with the credible interval.
- Sometimes this can be done analytically.

Highest Posterior Density Intervals and Sets, Example

HPD Interval for a Different Dataset



Multivariate Linear Modeling in Matrix Notation

- We have not yet described models in matrix algebra form, so rewrite the varying intercept/varying slope specification as:

$$y_i \sim \mathcal{N}(\mathbf{X}_i \mathbf{B}_{j[i]}, \sigma_y^2), \qquad \text{for } i = 1, \ldots, n$$

$$\mathbf{B}_j \sim \mathcal{N}(\mathbf{M}_B, \mathbf{\Sigma}_B), \qquad \text{for } j = 1, \ldots, J$$

  where:

  - $\mathbf{X}$ is the 919×2 matrix of explanatory variables with a leading column of 1's and the second column for floor.
  - $\mathbf{B} = (\alpha, \beta)$ is the $2 \times J$ matrix of estimated coefficients for the J groups from $\hat{\alpha}_j = \gamma_0^{\alpha} + \gamma_1^{\alpha} u_j$ and $\hat{\beta}_j = \gamma_0^{\beta} + \gamma_1^{\beta} u_j$.
  - $\mathbf{M}_B = (\mu_\alpha, \mu_\beta)$ is the 2-length mean vector of the $\mathbf{B}$ values.
  - $\mathbf{\Sigma}_B$ is the 2×2 variance-covariance matrix of the $\mathbf{B}$ values.

Multivariate Linear Modeling in Matrix Notation

- This is the matrix-notated version of the model that started Chapter 13.
- First level:

$$y_i \sim N(\alpha_{j[i]} + \mathbf{X}_i\beta_{j[i]}, \sigma_y^2), \qquad i = 1, \ldots, n$$

- Second level:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right) \qquad j = 1, \ldots, J$$

- We can also differentiate between modeled and un-modeled coefficients:

$$y_i \sim N(\mathbf{X}_i^0\boldsymbol{\beta}^0 + \mathbf{X}_i\boldsymbol{\beta}_{j[i]}, \sigma_y^2), \qquad i = 1, \ldots, n$$

where $\mathbf{X}_i^0$ denotes the explanatory variables for the ith case that are not given a hierarchy.

Multivariate Linear Modeling in Matrix Notation

- We have already estimated this model:

$$\hat{\mathbf{M}}_B = (1.4628, -0.6811)$$

$$\hat{\mathbf{\Sigma}}_B = \begin{pmatrix} \hat{\sigma}_a^2 & \hat{\rho}\hat{\sigma}_a\hat{\sigma}_b \\ \hat{\rho}\hat{\sigma}_a\hat{\sigma}_b & \hat{\sigma}_b^2 \end{pmatrix} = \begin{pmatrix} 0.122 & -0.337 \times 0.349 \times 0.344 \\ -0.337 \times 0.349 \times 0.344 & 0.188 \end{pmatrix}$$

since:

```
M3 <- lmer(y ~ 1 + x + (1 + x | county)); summary(M3)
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 county   (Intercept) 0.122    0.349
          x           0.118    0.344    -0.337
 Residual             0.557    0.746
number of obs: 919, groups: county, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)   1.4628     0.0539   27.15
x            -0.6811     0.0876   -7.78
```

Multivariate Linear Modeling in Matrix Notation

- Generalize this to have: J groups, K individual-level explanatory variables, and L group-level explanatory variables.
- So the last model had $K = L = 2$ with floor as the individual-level explanatory variable and uranium as the group-level explanatory variable, each level also having an intercept.
- The generalized specification is therefore:

$$y_i \sim \mathcal{N}(\underset{(1\times K)}{\mathbf{B}_{j[i]}}\ \underset{(K\times 1)}{\mathbf{X}_i}\ , \sigma_y^2), \qquad \text{for } i = 1, \ldots, n$$

$$\underset{(1\times K)}{\mathbf{B}_j}\ \sim \mathcal{N}(\ \underset{(1\times L)}{\mathbf{U}_j}\ \ \underset{(L\times K)}{\mathbf{G}}\ ,\ \underset{(K\times K)}{\mathbf{\Sigma}_B}\ ), \qquad \text{for } j = 1, \ldots, J$$

where:

  - ▷ $\mathbf{B}$ is the $J \times K$ matrix of individual-level coefficients
  - ▷ $\mathbf{U}$ is the $J \times L$ matrix of group-level explanatory variables
  - ▷ $\mathbf{G}$ is the $L \times K$ matrix of group-level coefficients.

Multivariate Linear Modeling in Matrix Notation

- Return to the model:

```
M4 <- lmer(y ~ x + u.full + x:u.full + (1 + x | county)); summary(M4)
Random effects:
Groups   Name        Variance Std.Dev. Corr
 county   (Intercept) 0.0156   0.125
 x           0.0941   0.307    0.409    0.409
 Residual             0.5617   0.749
number of obs: 919, groups: county, 85

Fixed effects:
            Estimate Std. Error t value      Correlation of Fixed Effects:
(Intercept)   1.4686     0.0353    41.6              (Intr) x      u.full
x            -0.6710     0.0844    -7.9      x        -0.241
u.full        0.8081     0.0907     8.9      u.full    0.207 -0.092
x:u.full     -0.4195     0.2271    -1.8      x:u.full -0.093  0.173 -0.231
```

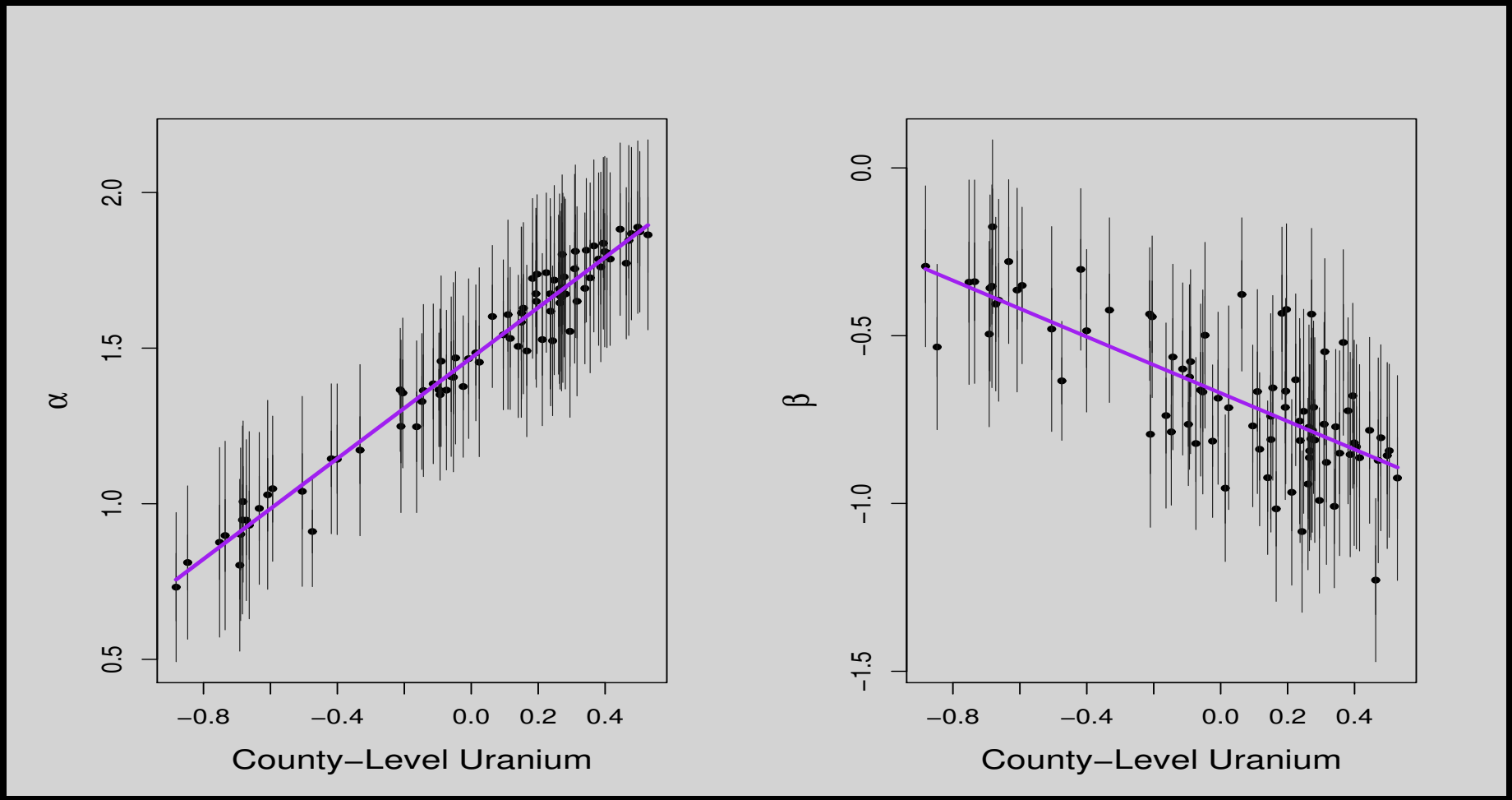
- Such that with $K = L = 2, J = 85$:
  - $\underset{(85\times 2)}{\mathbf{U}}$ is `[1,x]` for each county on each row.
  - $\underset{(2\times 2)}{\mathbf{G}}$ is:

$$\begin{bmatrix} \gamma_0^{\alpha} & \gamma_0^{\beta} \\ \gamma_1^{\alpha} & \gamma_1^{\beta} \end{bmatrix} = \begin{bmatrix} 1.4686 & -0.6710 \\ 0.8081 & -0.4195 \end{bmatrix}$$

  - $\underset{(85\times 2)}{\mathbf{B}} = \underset{(85\times 2)(2\times 2)}{\hat{\mathbf{U}}\ \mathbf{G}}$ is estimated by:

```
(a.hat.M4 <- coef(M4)$county[,1] + coef(M4)$county[,3]*u)
(b.hat.M4 <- coef(M4)$county[,2] + coef(M4)$county[,4]*u)
```

Figure 13.2 from Gelman & Hill



Modeling the Variance-Covariance Matrix

- When the number of modeled explanatory variables (L) is more than 2, then modeling the correlation parameters (ρ) is more challenging.
- Each correlation is constrained to $[-1 : 1]$, and the $\underset{(K\times K)}{\boldsymbol{\Sigma}_B}$ matrix must be positive definite: $\mathbf{q}'\mathbf{X}\mathbf{q} > 0$ for any conformable, non-null $\mathbf{q}$.
- This means that off-diagonal values are not independent and have constraints imposed by secondary relationships, for example: $r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1$.
- G&H give the example: $r_{12} = 0.9, r_{13} = 0.9$, then $r_{23} \geq 0.62$ from:

$$0.9 + 0.9 + r_{23}^2 - 2(0.9)(0.9)r_{23} \leq 1$$

$$r_{23}^2 - 1.62r_{23} + 0.8 \leq 0$$

- This is called a *jointly constrained parameter space.*

Modeling the Variance-Covariance Matrix

- We want to *model* the variance-covariance matrix for the estimated parameters $\boldsymbol{\Sigma}_B$, with diagonal elements $\boldsymbol{\Sigma}_{kk} = \sigma_k^2$, and off-diagonal elements $\boldsymbol{\Sigma}_{kl} = \rho_{kl}\sigma_k\sigma_l$.
- Start with specifying a *Wishart* distribution for the now random quantity $\boldsymbol{\Sigma}$:
  - PDF: $\mathcal{W}(\boldsymbol{\Sigma}|\alpha, \boldsymbol{\beta}) = \frac{|\boldsymbol{\Sigma}|^{(\alpha-(k+1))/2}}{\Gamma_k(\alpha)|\boldsymbol{\beta}|^{\alpha/2}} \exp[-\text{tr}(\boldsymbol{\beta}^{-1}\boldsymbol{\Sigma})/2]$

    where: $\Gamma_k(\alpha) = 2^{\alpha k/2}\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{\alpha+1-i}{2}\right),\ 2\alpha > k-1,$

    $\boldsymbol{\beta}$ symmetric nonsingular (full rank), and $\boldsymbol{\Sigma}$ symmetric positive definite.
  - $E[\boldsymbol{\Sigma}_{ij}] = \alpha\boldsymbol{\beta}_{ij}$
  - $\text{Var}[SI_{ij}] = \alpha(\boldsymbol{\beta}_{ij}^2 + \boldsymbol{\beta}_{ii}\boldsymbol{\beta}_{jj})$
  - $\text{Cov}[SI_{ij}, \boldsymbol{\Sigma}_{kl}] = \alpha(\boldsymbol{\beta}_{ik}\boldsymbol{\beta}_{jl} + \boldsymbol{\beta}_{il}\boldsymbol{\beta}_{jk})$.
- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ giving a $p \times n$ data matrix $\mathbf{X}$. The distribution of $M = \mathbf{X}\mathbf{X}'$ is a Wishart distribution.
- To make this application simple, define the scale $\boldsymbol{\beta}$ to be a $K \times$K diagonal matrix, and degrees of freedom to be $\alpha = K + 1$ (the number of group-level explanatory variables plus 1).

Modeling the Variance-Covariance Matrix

- If $\Upsilon \sim \mathcal{W}(\alpha, \boldsymbol{\beta})$, then $\Upsilon^{-1} \sim \mathcal{IW}(\alpha, \boldsymbol{\beta})$.
- Setting the degrees-of-freedom parameter to $K+1$ has the effect of setting a uniform distribution on the individual correlation parameters in $[-1, 1]$.
- However, this also constrains the σ^2 parameters, which should be informed by the data alone.
- A solution to this problem is to specify a new vector of chosen scale parameters:

$$\boldsymbol{\Sigma}_B = \xi \mathbf{Q} \xi,$$

  where $\mathbf{Q} \sim \mathcal{IW}(\boldsymbol{\beta}, K+1)$.

- Now the diagonal values are unconstrained (except for being positive) unscaled covariance values multiplied by the scaling factors:

$$\sigma_k^2 = \boldsymbol{\Sigma}_{kk} = \xi_k^2 \mathbf{Q}_{kk}, \qquad \text{for } k = 1, \ldots, K.$$

- And the off-diagonal values are:

$$\boldsymbol{\Sigma}_{kl} = \xi_k \xi_l \mathbf{Q}_{kl}, \qquad \text{for } k = 1, \ldots, K \; l = 1, \ldots, L.$$

Modeling the Variance-Covariance Matrix

- Expressing this in terms of the standard deviations:

$$\sigma_k = \|\xi_k\| \sqrt{\mathbf{Q}_{kk}}$$

and correlations:

$$\rho_{kl} = \boldsymbol{\Sigma}_{kl} / (\sigma_k \sigma_l)$$

- Note that the parameters in ξ and $\mathbf{Q}$ are tied together.
- Now we have an intuitive way to express relationships within the hierarchy between explanatory variables.

Multilevel Modeling As an Alternative to Selecting Regression Predictors

- Witte et al. (1994, and p.294 of G&H), give a logistic model from a case-control study of 362 people where the outcome variable is cancer incidence or not, using consumption information on 87 foods and 5 background variables.
- Challenge: regressing cancer $(0, 1)$ on 362×87 matrix $\mathbf{X}$, and a $362 \times (1 + 5)$ matrix $\mathbf{X}^0$.
- Fortunately we also have the 87 foods broken down in to levels of 35 nutrients expressed in the 87×35 matrix $\mathbf{Z}$.
- A multilevel model that increases the degrees of freedom by reducing food parameters from 87 to 35:

$$p(y_i = 1) = \text{logit}^{-1}(\mathbf{X}_i^0 \boldsymbol{\beta}^0 + \mathbf{X}_i \boldsymbol{\beta}_{j[i]}), \qquad i = 1, \ldots, 362$$

$$\boldsymbol{\beta}_j = N(\mathbf{Z}_j \boldsymbol{\gamma}, \sigma_{\boldsymbol{\beta}}^2), \qquad j = 1, \ldots, 87.$$

- So $\mathbf{Z}$ gives predictive effects for foods controlling for nutrients, where:
  - $\sigma_{\boldsymbol{\beta}}^2 \to 0$ means that all food variation explained by nutrients.
  - $\sigma_{\boldsymbol{\beta}}^2 \to \infty$ means that the food-nutrient relationship is so varied that it does not help at all.

Survival Models

- The *survival function* gives the probability of surviving past time t:

$$S(t) = p(T \geq t), \quad t \geq 0$$

  where T is the random variable for life length, and t is a fixed point of interest.

- In basic models, $S(t)$ is assumed to be smooth and uniformly differentiable at all points.

- Notation for the time of interest is either t or t_0.

- Places to get free high-quality introductions to survival models:
  `http://www.amstat.org/chapters/northeasternillinois/pastevents/presentations/summer05_Ibrahim_J.pdf`
  `http://biostat.mc.vanderbilt.edu/wiki/pub/Main/CourseBios312/survivalintro.pdf`
  `https://perswww.kuleuven.be/~u0018341/documents/survival.pdf`
  `http://anson.ucdavis.edu/ hiwang/teaching/10fall/R_tutorial%201.pdf`

Density Function

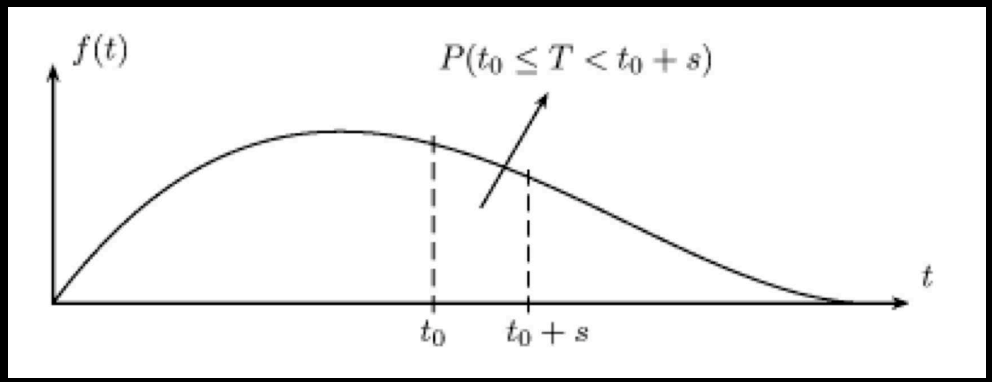
- The density function is related to the survival by:

$$f(t) = -\frac{\partial}{\partial t} S(t), \quad t \geq 0.$$

- For a very small s value and arbitrary t_0:

$$p(t_0 \leq T < t_0 + s) \approx s f(t_0),$$

which is illustrated by the figure at right as an approximation since this is not a rectangle.



- More exactly:

$$f(t) = \lim_{s \to 0} \frac{p(t \leq T < t + s)}{s}, \quad t \geq 0,$$

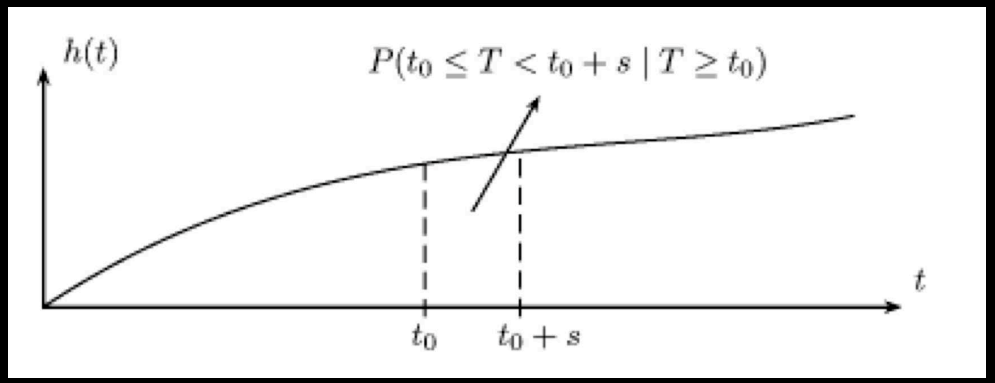
which is an unconditional statement.

Hazard Function

- The hazard function is also called the *instantaneous hazard rate*, the *instantaneous death rate*, the *intensity rate*, and the *force of mortality*.
- This is the instantaneous probability of the event at exactly t, given no event before then.
- Consider the s getting very small, then:

$$h(t_0) = \lim_{s \to 0} \frac{p(t_0 \leq T < t_0 + s | T \geq t_0)}{s},$$

(for $t_0 \geq 0$) which is a conditional statement.



The Cumulative Hazard Function

- The cumulative hazard function is just the integral of the hazard function (like a CDF):

$$F(t) = H(t) = \int_0^t h(x)dx, \quad t \geq 0.$$

- Here x is just an integration variable (it goes away).
- This form successively accumulates risk as time continues.
- The cumulative hazard function is easier to manipulate in statistical models than the hazard function, which is a property we will exploit.

Proportional Hazards

- If $h_0(t)$ and $h_1(t)$ are hazard functions from two separate distributions, they are *proportional* if:

$$h_1(t) = \phi h_0(t), \quad \forall t \geq 0$$

  where the scalar $\phi > 0$.

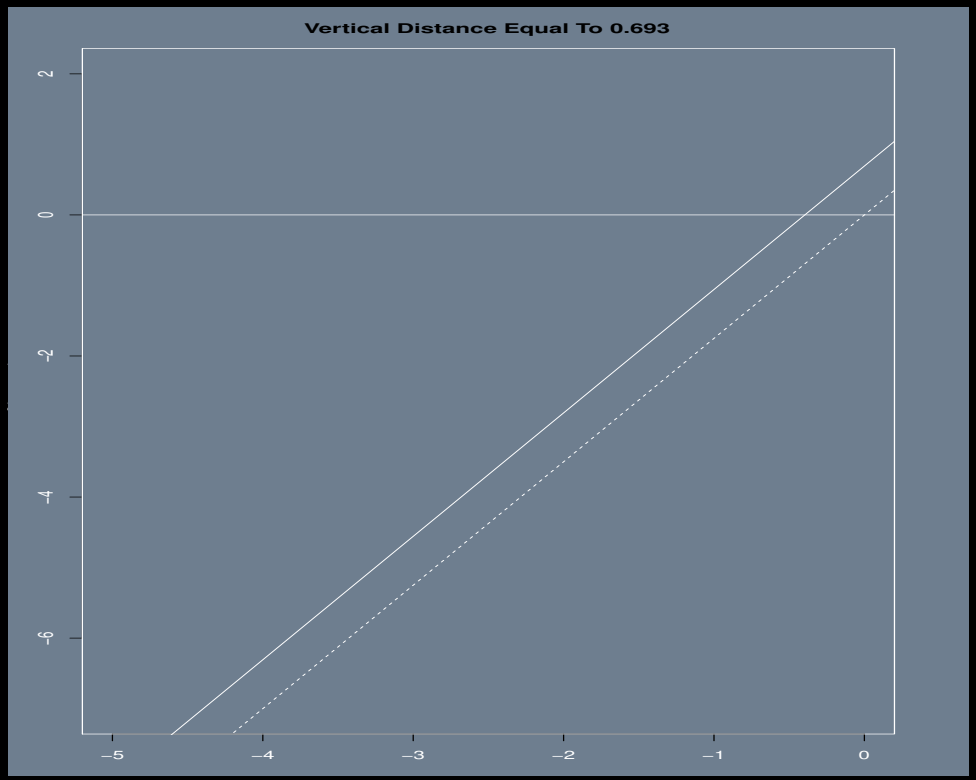
- This property caries over to the corresponding cumulative hazard functions:

$$H_1(t) = \phi H_0(t), \quad \forall t \geq 0$$

- Note that that ϕ is constant and therefore does not depend on t (eg. women have a survival advantage at all ages).

More on Proportional Hazards

- The proportional hazards assumption is important for Cox models.
- One way to check is a log-log plot: time versus hazard, also called a *Weibull plot*:
- For example with $\phi = 2$ then the vertical distance is $\log(2) = 0.693$:



A Note On Proportional Hazards

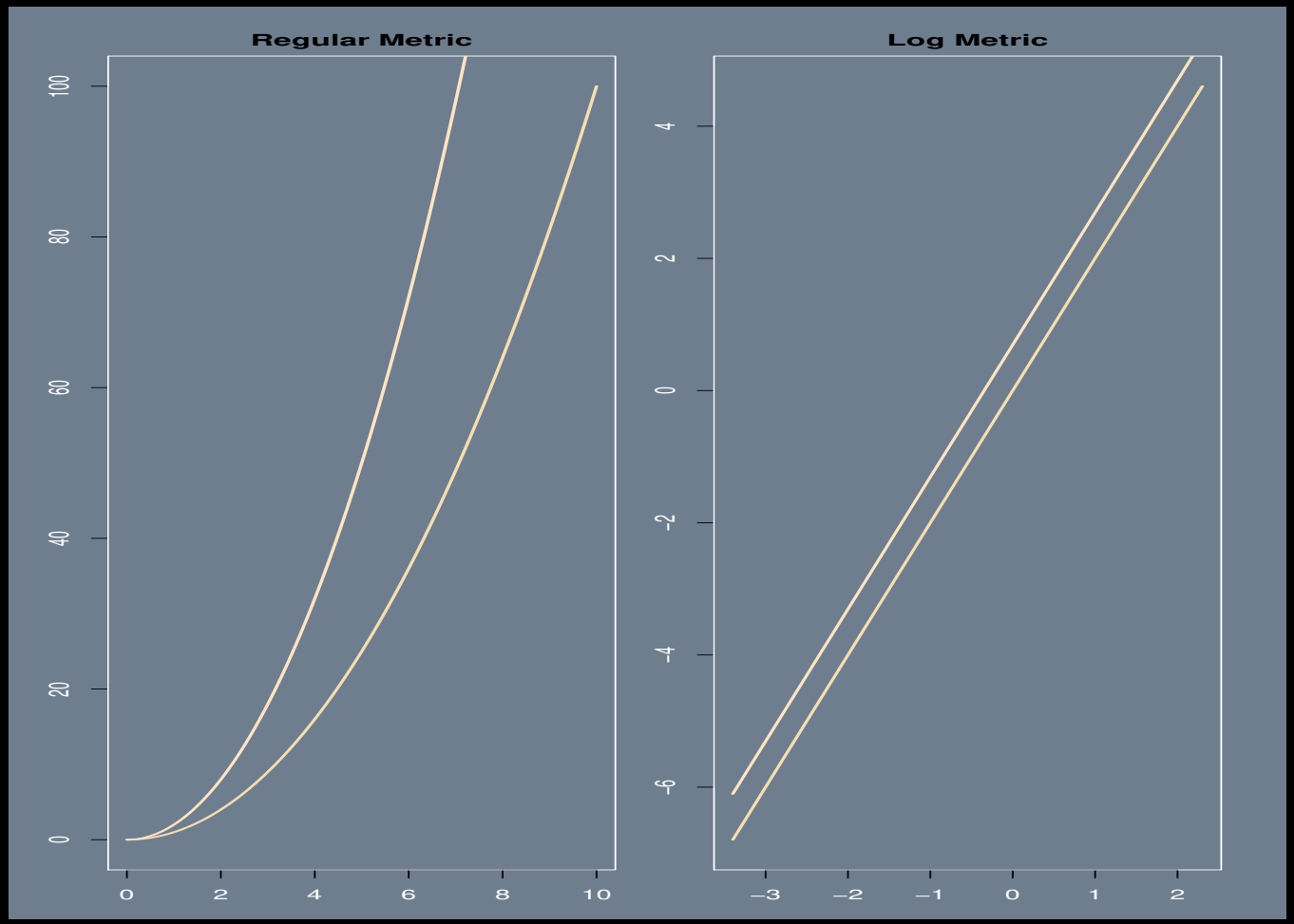
- To increase our intuition, consider the simple functions:

$$h_0(t) = t^2, \quad h_1(t) = 2h_0(t) = 2t^2, \quad t \geq 0.$$

- Let's plot these on the regular and log scale:

```
par(oma=c(1,1,1,1),mar=c(2,2,2,1),mfrow=c(1,2),col.axis="white",
    col.lab="white",col.sub="white",col="white", bg="slategray")
dur <- seq(0,10,length=300)
plot(dur,dur^2,type="l",lwd=2,col="wheat", main="Regular Metric")
lines(dur,2*dur^2,lwd=2,col="bisque")
plot(log(dur),log(dur^2),type="l",lwd=2,col="wheat", main="Log Metric")
lines(log(dur),log(2*dur^2),lwd=2,col="bisque")
```

A Note On Proportional Hazards



General Proportional Hazards Model

- This provides the specification:

$$h(t;\mathbf{x}) = h_0(t)\prod_{\ell=1}^{k}\exp(x_1\beta_1 + x_1\beta_2 + \cdots + x_{k-1}\beta_{k-1} + x_k\beta_k) = h_0(t)\exp(\mathbf{x}\boldsymbol{\beta}).$$

  (notice the lack of an intercept).

- We want to allow $\mathbf{x}$ to take on any values, not just binary indicators, as in standard regression.
- For $i = 1, \ldots, n$ cases in the data, define the ith data element as $(t_{i0}, t_i, d_i, \mathbf{x}_i)$, where:
  - t_{i0} is a left truncation time: if $y_{i0} = 0,\ \forall i$, then drop this
  - t_i is end time
  - d_i is event indicator: 1 if TRUE, 0 if FALSE (end of study or right-censoring).
  - $\mathbf{x}_i$ is a vector of explanatory variables.
- The *survival object* is then (t_{i0}, t_i, d_i).
- So now we need to estimate $h_0(t)$ and $\boldsymbol{\beta}$.

Estimation of the Continuous Baseline Cumulative Hazard

- First consider the interval to be summed: $j : t_j < t$, which is all of the time periods that come before j,
- Define R_j to be the number at risk at time period j.
- We say that $\ell \in R_j$ if the ℓth case (out of the n possible) is in the risk group at time j.
- $\hat{\boldsymbol{\beta}}$ is the estimated coefficient vector from MLE.
- $\mathbf{x}_i$ is the ith cases vector of explanatory variables, continuous or discrete.
- The standard estimate is:

$$\hat{H}_0(t) = \sum_{j:t_j<t} \frac{d_i}{\sum_{\ell \in R_j} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}})}$$

Continuous Time, Factor Covariate Interpretation

- For individual i at time t with explanatory variable vector $\mathbf{x}_i$ the hazard function is:

$$\hat{h}(t;\mathbf{x}_i) = \hat{h}_0(t)\exp(\mathbf{x}_i\boldsymbol{\beta})$$

  which shows the porportionality of the regression information relative to the baseline:

$$\frac{\hat{h}(t;\mathbf{x}_i)}{\hat{h}_0(t)} = \exp(\mathbf{x}_i\boldsymbol{\beta})$$

  (relative risk to the baseline), and also

$$\log\left(\frac{\hat{h}(t;\mathbf{x}_i)}{\hat{h}_0(t)}\right) = \mathbf{x}_i\boldsymbol{\beta}$$

  (log relative risk to the baseline).

- $\hat{h}(t;\mathbf{x}_i)$ allows calculation of the survival function for the ith case:

$$\hat{S}(t,\mathbf{x}_i) = \exp(-\hat{h}(t;\mathbf{x}_i)).$$

Continuous Time, Continuous Coefficient Interpretation

- Assume for the moment a single explanatory variable just for simplicity:

$$h(t;x) = h_0(t)\exp(x\boldsymbol{\beta})$$

- Consider the effect of adding 1 to this explanatory variable the way we often discuss with linear models:

$$\frac{h(t;x+1)}{h(t;x)} = \frac{h(t)\exp(\beta(x+1))}{h(t)\exp(\beta x)} = \frac{\exp(\beta(x+1))}{\exp(\beta x)} = \frac{\exp(\beta x)\exp(\beta)}{\exp(\beta x)} = \exp(\beta).$$

- So incrementing x by one increases the relative risk (hazard ratio) by $\exp(\beta)$, which easy to interprete.

- This is why `R` (and other packages) routinely provide the exponent of the coefficient as well in the output.

Proportional Hazards In Discrete Time

- Proportional hazards in discrete time is a set of conditional probabilities, which are by definition bounded by $[0:1]$.
- So ratios can give awkard numbers that exceed one.
- Fix: assume continuous time that is segmented into a set of k period by the mechanism of measurement: $0 = t_0 < t_1 < t_2 < \cdots < t_k = \infty$.
- Then for the random variable T we have:

$$p(t_i \leq T < t_i | T \geq t_{i-1}, \mathbf{x}) = \frac{S(t_{i-1}|\mathbf{x}) - S(t_i|\mathbf{x})}{S(t_{i-1}|\mathbf{x})}$$

Since survival functions work "backwards" relative to cumulative hazard functions: $S(t) = p(T \geq t) = 1 - H(t)$, and continuing...

$$\cdots = 1 - \frac{S(t_i|\mathbf{x})}{S(t_{i-1}|\mathbf{x})} = 1 - \left(\frac{S_0(t_i)}{S_0(t_{j-1})}\right)^{\exp(\mathbf{x}\boldsymbol{\beta})} = 1 - (1 - h_i)^{\exp(\mathbf{x}\boldsymbol{\beta})}$$

where:

$$h_i = p(t_{i-1} \leq T < t_i | T \geq t_{i-1}, \mathbf{x})$$

to give the definition of proportional hazards in discrete time.

Example with Kidney Data

Recurrence times to infection at point of insertion of the catheter for kidney patients using portable dialysis.

```
> library(survival)
> data(kidney)
> head(kidney,10)
    id time status age sex disease frail
1    1    8      1  28   1   Other   2.3
2    1   16      1  28   1   Other   2.3
3    2   23      1  48   2      GN   1.9
4    2   13      0  48   2      GN   1.9
5    3   22      1  32   1   Other   1.2
6    3   28      1  32   1   Other   1.2
7    4  447      1  31   2   Other   0.5
8    4  318      1  32   2   Other   0.5
9    5   30      1  10   1   Other   1.5
10   5   12      1  10   1   Other   1.5
```

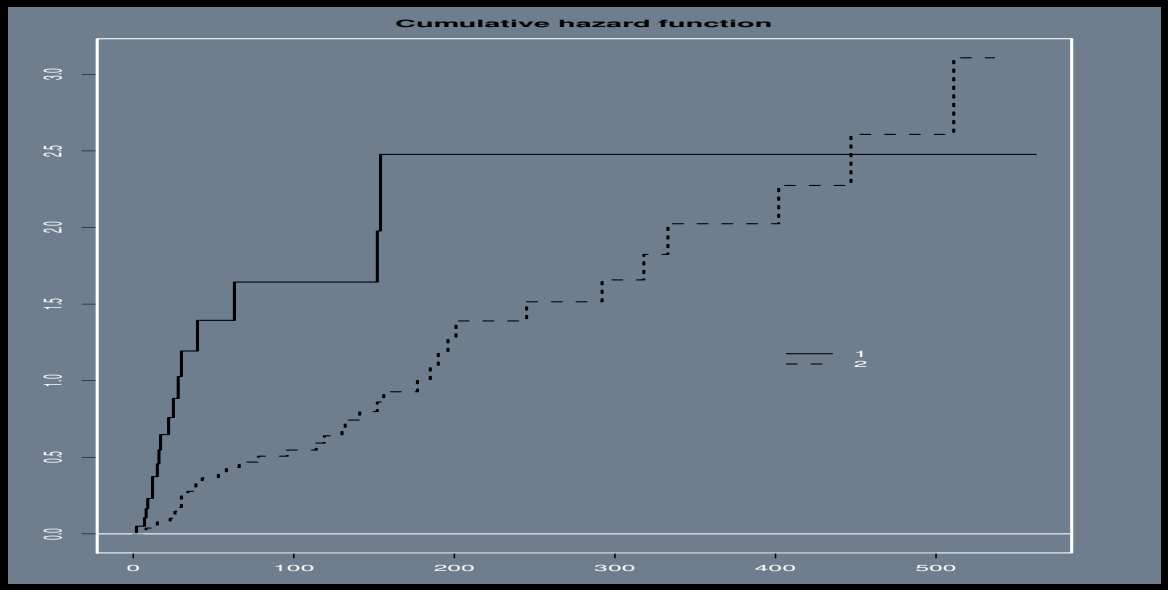
Example with Kidney Data

- CA McGilchrist, CW Aisbett (1991), Regression with frailty in survival analysis. Biometrics 47, 461–66.
- `patient: id`
- `time: time`
- `status: event status`
- `age: in years`
- `sex: 1=male, 2=female`
- `disease: disease type (0=GN, 1=AN, 2=PKD, 3=Other)`
- `frail: frailty estimate from original paper`
- Looking at dimensions:

```
dim(kidney)
[1] 76  7
length(unique(kidney$id))
[1] 38
```

Check Proportional Hazards Assumption

```
library(eha)
postscript("Class.Multilevel/Images/kidney.ph.ps")
par(mfrow=c(1,1),mar=c(3,3,3,3),col.axis="white",
    col.lab="white", col.sub="white",col="white",bg="slategray",lwd=2)
with(kidney, plot(Surv(rep(0,nrow(kidney)),time,status),strat=sex))
dev.off()
```



Cox Proportional Hazards in R

```
kidney1.out <- coxph(Surv(time,status) ~ sex + age, data=kidney)
summary(kidney1.out)
        coef exp(coef) se(coef)     z Pr(>|z|)
sex -0.82931   0.43635  0.29895 -2.77   0.0055
age  0.00203   1.00203  0.00925  0.22   0.8261

    exp(coef) exp(-coef) lower .95 upper .95
sex     0.436      2.292     0.243     0.784
age     1.002      0.998     0.984     1.020

Concordance= 0.662  (se = 0.046 )
Rsquare= 0.089   (max possible= 0.993 )
Likelihood ratio test= 7.12  on 2 df,   p=0.0285
Wald test            = 8.02  on 2 df,   p=0.0181
Score (logrank) test = 8.45  on 2 df,   p=0.0147
```

Note: Concordance is p(agreement) for any two randomly chosen observations, where in this case agreement means that the observation with the shorter survival time of the two also has the larger risk. For continuous covariates concordance is equivalent to Kendall's tau, which is (# concordant pairs - # discordant pairs)/$0.5n(n-1)$, and for logistic regression is is equivalent to the area under the ROC curve

Frailty Model

```
install.packages("coxme")
library(coxme)
data(kidney)
kidney2.out <- coxme(Surv(time,status) ~ sex + age + (1|id), data=kidney)
print(kidney2.out)
```

```
Cox mixed-effects model fit by maximum likelihood
   Data: kidney
   events, n = 58, 76
   Iterations= 6 34
                    NULL Integrated  Fitted
Log-likelihood -187.9     -181.9 -166.17
```

Notes: Three models fit with log-likelihood values. . . NULL model with no covariates, Integrated with zero variance random effects, and Fitted with non-zero variance random effects. Numbers after `events` and `Iterations` apply to Integrated log-likelihood and Fitted (Penalized) log-likelihood.

Frailty Model

```
                  Chisq    df          p   AIC    BIC
Integrated loglik 12.00  3.00 0.00739530  6.00  -0.18
Penalized loglik  43.48 14.75 0.00011458 13.97 -16.43

Model:  Surv(time, status) ~ sex + age + (1 | id)
Fixed coefficients
          coef exp(coef) se(coef)     z      p
sex -1.3549853   0.25795  0.41713 -3.25 0.0012
age  0.0042892   1.00430  0.01171  0.37 0.7100

Random effects
  Group Variable  Std Dev Variance
  id    Intercept 0.67545 0.45623
```

Frailty Model

```
fixef(kidney2.out)
          sex           age
-1.3549852701  0.0042892004
ranef(kidney2.out)
 0.514285501  0.312981298  0.184422411 -0.489289904  0.254271656  0.063924646
           7            8            9           10           11           12
 0.677538398 -0.367396556 -0.039733287 -0.421879955 -0.097571798  0.052933687
          13           14           15           16           17           18
 0.332254323 -0.410603097 -0.575683144  0.167737805 -0.138329832 -0.131842520
          19           20           21           22           23           24
-0.393831993  0.103431022 -1.547860215 -0.386972008  0.443174433  0.074684951
          25           26           27           28           29           30
 0.064824787 -0.327930094  0.095907561  0.465382028  0.405734032  0.309895414
          31           32           33           34           35           36
 0.417557869  0.211492863  0.204591001 -0.113169928  0.444184326 -0.203473269
          37           38
 0.143432010 -0.299074418
```

Two Non-nested hierarchies

```
kidney3.out <- coxme(Surv(time,status) ~ sex + age + (1|id) + (1|disease),
                data=kidney)
print(kidney3.out)

Cox mixed-effects model fit by maximum likelihood
   Data: kidney
   events, n = 58, 76
   Iterations= 10 54
                NULL Integrated  Fitted
Log-likelihood -187.9     -181.9 -166.14
```

Two Non-nested hierarchies

```
                  Chisq    df          p   AIC    BIC
Integrated loglik 12.00  4.00 0.01738300  4.00  -4.25
 Penalized loglik 43.52 14.77 0.00011408 13.97 -16.47

Model:  Surv(time, status) ~ sex + age + (1 | id) + (1 | disease)
Fixed coefficients
          coef exp(coef) se(coef)     z      p
sex -1.3559216   0.25771 0.417331 -3.25 0.0012
age  0.0042822   1.00429 0.011722  0.37 0.7100

Random effects
  Group   Variable  Std Dev    Variance
  id      Intercept 0.67606440 0.45706308
  disease Intercept 0.01998455 0.00039938
```