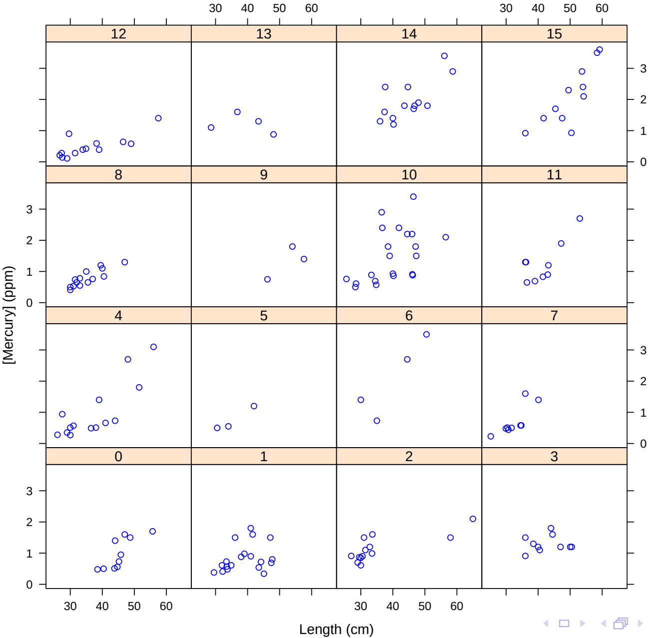
Hierarchical Regression Models

Hoff Chapter 11

November 12, 2015

NC Mercury in Fish data



Models

Consider the following models for $\log \text{MERCURY}$ as a function of $\log \text{LENGTH}$:

Models

Consider the following models for $\log \text{MERCURY}$ as a function of $\log \text{LENGTH}$:

1. $\log \text{MERCURY}_{ij} = \beta_0 + \beta_1 \log \text{LENGTH}_{ij}$ (common line for all stations)

Models

Consider the following models for $\log \text{MERCURY}$ as a function of $\log \text{LENGTH}$:

1. $\log \text{MERCURY}_{ij} = \beta_0 + \beta_1 \log \text{LENGTH}_{ij}$ (common line for all stations)
2. $\log \text{MERCURY}_{ij} = \beta_{0j} + \beta_1 \log \text{LENGTH}_{ij}$ (parallel regression lines)

Models

Consider the following models for $\log \text{MERCURY}$ as a function of $\log \text{LENGTH}$:

1. $\log \text{MERCURY}_{ij} = \beta_0 + \beta_1 \log \text{LENGTH}_{ij}$ (common line for all stations)
2. $\log \text{MERCURY}_{ij} = \beta_{0j} + \beta_1 \log \text{LENGTH}_{ij}$ (parallel regression lines)
3. $\log \text{MERCURY}_{ij} = \beta_{0j} + \beta_{1j} \log \text{LENGTH}_{ij}$ (separate lines for each station)

Models

Consider the following models for $\log \text{MERCURY}$ as a function of $\log \text{LENGTH}$:

1. $\log \text{MERCURY}_{ij} = \beta_0 + \beta_1 \log \text{LENGTH}_{ij}$ (common line for all stations)
2. $\log \text{MERCURY}_{ij} = \beta_{0j} + \beta_1 \log \text{LENGTH}_{ij}$ (parallel regression lines)
3. $\log \text{MERCURY}_{ij} = \beta_{0j} + \beta_{1j} \log \text{LENGTH}_{ij}$ (separate lines for each station)

Use ANOVA to compare the 3 models

Fitting Models with Categorical Predictors in R

```
fish$S = factor(fish$STATION) # convert to categorical
fish.com = lm(log(MERCURY) ~ 1 + log(LENGTH), data=fish)
fish.par = lm(log(MERCURY) ~ S + log(LENGTH), data=fish)
fish.dif = lm(log(MERCURY) ~ S*log(LENGTH), data=fish)

anova(fish.com, fish.par, fish.dif)
Analysis of Variance Table

Model 1: log(MERCURY) ~ 1 + log(LENGTH)
Model 2: log(MERCURY) ~ S + log(LENGTH)
Model 3: log(MERCURY) ~ S * log(LENGTH)
  Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
1    169 41.621
2    154 23.974  15    17.648 8.1515 5.051e-13 ***
3    139 20.062  15     3.912 1.8070   0.03918 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Model Comparison: ANOVA Extra Sum-of-Squares F test

Under all models the MSE = RSS/df under model (3) is an estimate of σ^2. If one of the simpler models is true (say model (2)), then we expect that the difference in RSS between model (2) and (3) will be small relative to the extra degrees of freedom.

Model Comparison: ANOVA Extra Sum-of-Squares F test

Under all models the MSE $=$ RSS/df under model (3) is an estimate of σ^2. If one of the simpler models is true (say model (2)), then we expect that the difference in RSS between model (2) and (3) will be small relative to the extra degrees of freedom. Distribution of

$$F_{obs} = \frac{\frac{RSS_{H_2} - RSS_{H_3}}{df_{H_2} - df_{H_3}}}{\hat{\sigma}^2_{H_3}} = \frac{\frac{\Delta RSS}{\Delta df}}{\hat{\sigma}^2}$$

is an $F(df_{H_2} - df_{H_3}, df_{H_3})$; compare the p-value $= P(F > F_{obs})$

Model Comparison: ANOVA Extra Sum-of-Squares F test

Under all models the MSE $=$ RSS/df under model (3) is an estimate of σ^2. If one of the simpler models is true (say model (2)), then we expect that the difference in RSS between model (2) and (3) will be small relative to the extra degrees of freedom.
Distribution of

$$F_{obs} = \frac{\frac{RSS_{H_2} - RSS_{H_3}}{df_{H_2} - df_{H_3}}}{\hat{\sigma}^2_{H_3}} = \frac{\frac{\Delta RSS}{\Delta df}}{\hat{\sigma}^2}$$

is an $F(df_{H_2} - df_{H_3}, df_{H_3})$; compare the p-value $= P(F > F_{obs})$
The ANOVA output suggests that we would

- reject model (2) in favor of (3) at the $\alpha = 0.05$ level, but not at the 0.01 level

Model Comparison: ANOVA Extra Sum-of-Squares F test

Under all models the MSE = RSS/df under model (3) is an estimate of σ^2. If one of the simpler models is true (say model (2)), then we expect that the difference in RSS between model (2) and (3) will be small relative to the extra degrees of freedom. Distribution of

$$F_{obs} = \frac{\frac{RSS_{H_2} - RSS_{H_3}}{df_{H_2} - df_{H_3}}}{\hat{\sigma}^2_{H_3}} = \frac{\frac{\Delta RSS}{\Delta df}}{\hat{\sigma}^2}$$

is an $F(df_{H_2} - df_{H_3}, df_{H_3})$; compare the p-value $= P(F > F_{obs})$
The ANOVA output suggests that we would

- reject model (2) in favor of (3) at the $\alpha = 0.05$ level, but not at the 0.01 level
- reject model (1) at any $\alpha > 5.05 \times 10^{-13}$

Goals of Model

- Model (3) leads to a separate line for each of the 16 stations

Goals of Model

- Model (3) leads to a separate line for each of the 16 stations
- Model (2) leads to a separate intercept for each of the 16 stations

Goals of Model

- Model (3) leads to a separate line for each of the 16 stations
- Model (2) leads to a separate intercept for each of the 16 stations

How can we use these models to predict MERCURY levels for other locations?

Goals of Model

- Model (3) leads to a separate line for each of the 16 stations
- Model (2) leads to a separate intercept for each of the 16 stations

How can we use these models to predict MERCURY levels for other locations? View STATION as a random sample of locations

and build a hierarchical model

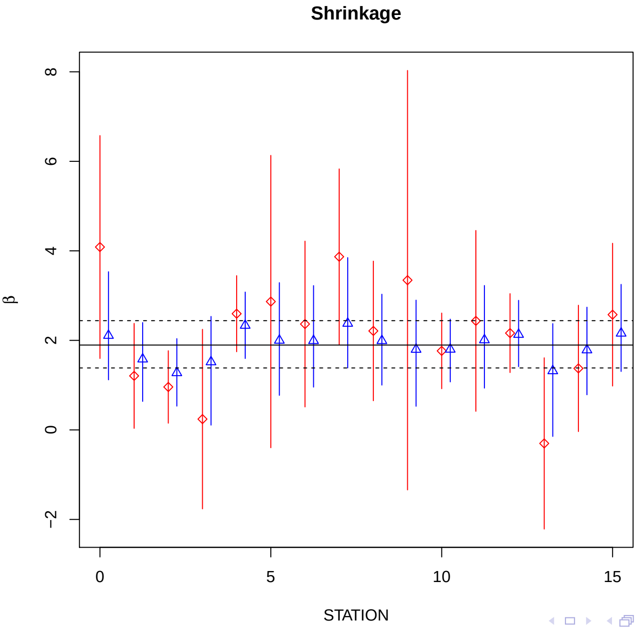
JAGS/BUGS Model - Non-Centered

```
for (n in 1:N){
 muj[n] <- alpha[station[n]]+beta[station[n]]*X[n]
 Y[n] ~ dnorm(muj[n], phi)
}

for (j in 1:J) {
  alpha[j] ~ dnorm(alpha.mu, alpha.phi)
  beta[j] ~ dnorm(beta.mu,  beta.phi)
}
phi ~ dgamma(.001, .001)
alpha.mu ~ dnorm(0.0, 1.0E-6)
alpha.sigma ~ dunif(0, 100)
alpha.phi <-1/(alpha.sigma*alpha.sigma)
beta.mu ~ dnorm(0.0, 1.0E-6)
beta.phi <- pow(beta.sigma, -2)
beta.sigma ~ dunif(0, 100)
```

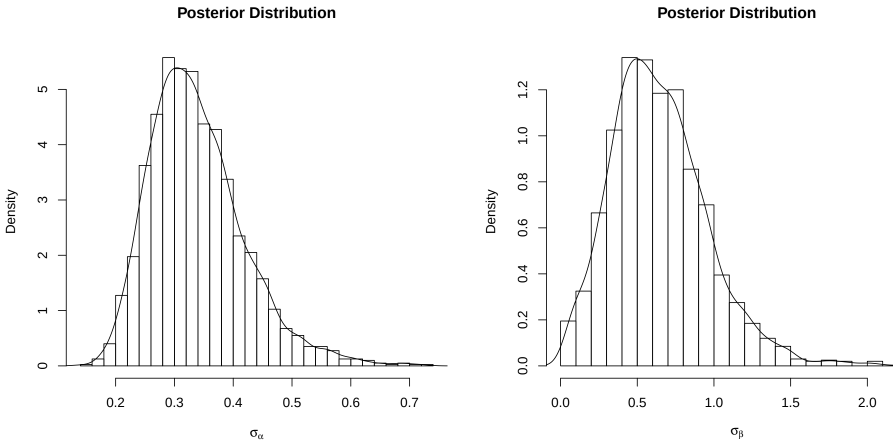
JAGS/BUGS Model - Centered

```
for (n in 1:N){
 muj[n] <- alpha[station[n]]+beta[station[n]]*(X[n]-xbar)
 Y[n] ~ dnorm(muj[n], phi)
}

for (j in 1:J) {
  alpha[j] ~ dnorm(alpha.mu, alpha.phi)
  beta[j] ~ dnorm(beta.mu,  beta.phi)
}
phi ~ dgamma(.001, .001)
alpha.mu ~ dnorm(0.0, 1.0E-6)
alpha.sigma ~ dunif(0, 100)
alpha.phi <-1/(alpha.sigma*alpha.sigma)
beta.mu ~ dnorm(0.0, 1.0E-6)
beta.phi <- pow(beta.sigma, -2)
beta.sigma ~ dunif(0, 100)
```

Shrinkage

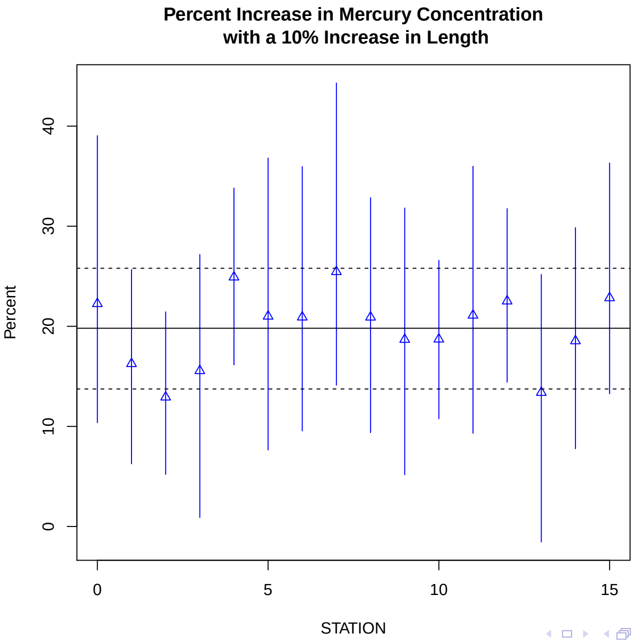


Variance Components



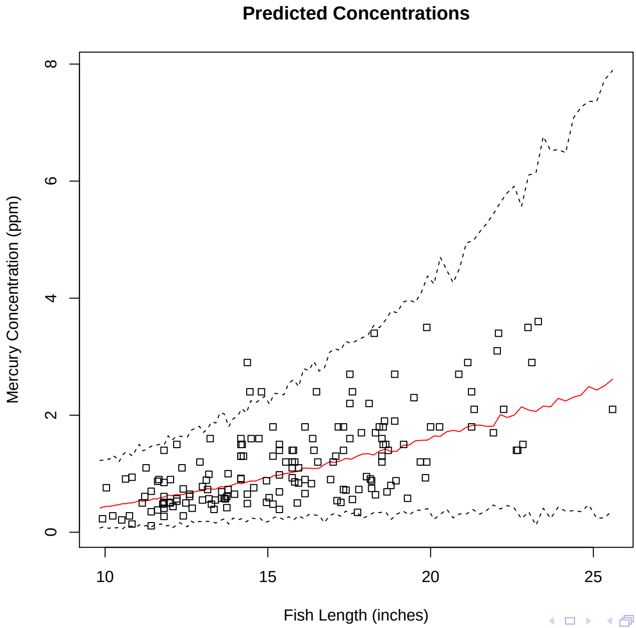
Interpretation of Coefficients

Percent Increase

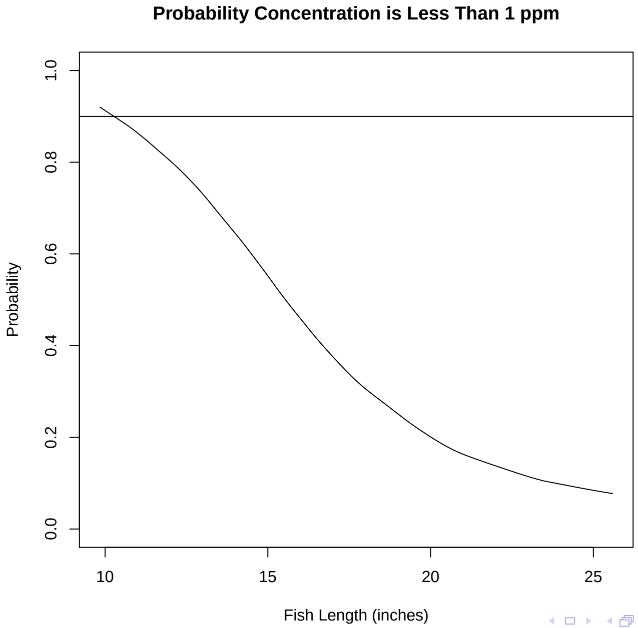


Predicting at a New Location

Predictions for a Random Location



Probability Concentration Does not Exceed 1ppm



Model Extensions

- Should intercepts or slopes depend on River? Add another level to the hierarchy.

Model Extensions

- Should intercepts or slopes depend on River? Add another level to the hierarchy.
- Sensitivity of results to prior on variance components?

Model Extensions

- Should intercepts or slopes depend on River? Add another level to the hierarchy.
- Sensitivity of results to prior on variance components? Alternative prior distribution for standard deviation is a half-Cauchy. (a Cauchy distribution restricted to $(0, \infty)$.

Model Extensions

- Should intercepts or slopes depend on River? Add another level to the hierarchy.
- Sensitivity of results to prior on variance components? Alternative prior distribution for standard deviation is a half-Cauchy. (a Cauchy distribution restricted to $(0, \infty)$.
- Scaled Beta2 prior?
- Inclusion of weight? Measurement error model for weight as a function of length?

Model Extensions

- Should intercepts or slopes depend on River? Add another level to the hierarchy.
- Sensitivity of results to prior on variance components? Alternative prior distribution for standard deviation is a half-Cauchy. (a Cauchy distribution restricted to $(0, \infty)$.
- Scaled Beta2 prior?
- Inclusion of weight? Measurement error model for weight as a function of length?