Linear Models

- Variety of linear models
- MLE derivation of parameters and se's [ref]
- Comparison to Bayesian
- Assumptions of linear models
- Relaxing these assumptions
Linear models

- Statistically, a model is judged based on whether it is linear or not with respect to the parameters.

\[
\begin{align*}
y &= \beta_0 + \beta_1 x_1 \\
y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 \\
y &= \beta_0 + \beta_1 x + \beta_2 x^2 \\
y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \cdot x_2 \\
y &= \beta_0 + \beta_1 \ln(x_1) + \beta_2 \exp(x_2) \\
y &= \beta_0 + \beta_1 I(TRT1) + \beta_2 I(TRT2)
\end{align*}
\]
Recall for simple linear model

\[ y = \beta_0 + \beta_1 x + \epsilon \]

\[ \beta_1 = \frac{\text{cov}[x, y]}{\text{var}[x]} \]

\[ \beta_0 = \bar{y} - \beta_1 \bar{x} \]

\[ \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \]
Parameter CI by Fisher Information

\[\ln L = -\frac{n}{2} \ln (2 \pi \sigma^2) - \sum \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2 \sigma^2}\]

\[
\frac{\partial \ln L}{\partial \beta_0} = \frac{1}{\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i) = \frac{1}{\sigma^2} \left[ \sum y_i - n \beta_0 - \beta_1 \sum x_i \right]
\]

\[
\frac{\partial \ln L}{\partial \beta_1} = \frac{1}{\sigma^2} \sum x_i (y_i - \beta_0 - \beta_1 x_i)
= \frac{1}{\sigma^2} \left[ \sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 \right]
\]
\[
\frac{\partial \ln L}{\partial \beta_0} = \frac{1}{\sigma^2} \left[ \sum y_i - n \beta_0 - \beta_1 \sum x_i \right]
\]

\[
\frac{\partial^2 \ln L}{\partial \beta_0^2} = - \frac{1}{\sigma^2} \left[ n + \frac{\partial \beta_1}{\partial \beta_0} \sum x_i \right]
\]

\[
\frac{\partial^2 \ln L}{\partial \beta_0^2} = - \frac{n}{\sigma^2} \left[ 1 - \frac{\bar{x}^2}{x^2} \right]
\]

\[
\frac{\partial^2 \ln L}{\partial \beta_0^2} = - \frac{n}{\sigma^2} \left[ \frac{x^2 - \bar{x}^2}{\bar{x}^2} \right]
\]

\[
\frac{\partial^2 \ln L}{\partial \beta_0^2} = -n \text{var} \left[ x \right]
\]

\[
\frac{\partial^2 \ln L}{\partial \beta_0^2} = -\frac{n \text{var} \left[ x \right]}{\sigma^2 x^2}
\]

From our MLE estimator,

\[
\beta_1 = \frac{\bar{x}y - \beta_0 \bar{x}}{x^2}
\]

\[
\frac{\partial \beta_1}{\partial \beta_0} = -\frac{\bar{x}}{x^2}
\]

\[
se_{\beta_0} = \frac{1}{\sqrt{I_{\beta_0}}}
\]

\[
se_{\beta_0} = \sigma \sqrt{\frac{x^2}{n \text{var} \left[ x \right]}}
\]
\[
\frac{\partial \ln L}{\partial \beta_1} = \frac{1}{\sigma^2} \left[ \sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 \right]
\]

\[
\frac{\partial^2 \ln L}{\partial \beta_1^2} = \frac{1}{\sigma^2} \left[ -\frac{\partial \beta_0}{\partial \beta_1} \sum x_i - \sum x_i^2 \right]
\]

\[
\frac{\partial^2 \ln L}{\partial \beta_0^2} = \frac{1}{\sigma^2} \left[ \bar{x} \sum x_i - \sum x_i^2 \right]
\]

\[
\frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_0} = -\frac{n}{\sigma^2} \text{var} [x]
\]

\[
se_{\beta_1} = \frac{\sigma}{\sqrt{n \text{var} [x]}}
\]
Multiple Regression via MLE

- Recall from our Bayesian derivation that we can express the regression likelihood in matrix form

\[
\tilde{y} | \beta, \sigma^2 \sim N ( \boldsymbol{X} \beta, \sigma^2 )
\]

\[
L \propto \sigma^{-n} \exp \left[ - \frac{(\tilde{y} - \boldsymbol{X} \beta)^T (\tilde{y} - \boldsymbol{X} \beta)}{2 \sigma^2} \right]
\]

\[
\ln L \propto -n \ln(\sigma) - \frac{(\tilde{y} - \boldsymbol{X} \beta)^T (\tilde{y} - \boldsymbol{X} \beta)}{2 \sigma^2}
\]

\[
\ln L \propto -n \ln(\sigma) - \frac{1}{2 \sigma^2} \left[ \tilde{y}^T \tilde{y} - y^T \boldsymbol{X} \beta - \beta^T \boldsymbol{X}^T \tilde{y} + \beta^T \boldsymbol{X}^T \boldsymbol{X} \beta \right]
\]
\[ \ln L \propto -n \ln(\sigma) - \frac{1}{2\sigma^2} \left[ y^T y - y^T X \beta - \beta^T X^T y + \beta^T X^T X \beta \right] \]

Vector derivative properties

\[
\frac{\partial A \beta}{\partial \beta} = \frac{\partial \beta^T A}{\partial \beta^T} = A \quad \frac{\partial \beta^T A \beta}{\partial \beta} = \beta^T A^T + \beta^T A
\]

\[
\begin{align*}
\frac{\partial \ln L}{\partial \beta} & \propto -\frac{1}{2\sigma^2} \left[ -2 y^T X + 2 \beta^T X^T X \right] = 0 \\
y^T X &= \beta^T X^T X \\
X^T y &= X^T X \beta
\end{align*}
\]

\[
\beta = (X^T X)^{-1} X^T y
\]

\[
\beta_1 = \frac{cov[x, y]}{var[x]}
\]
MLE vs Bayes

\[ \beta = (X^T X)^{-1} X^T y \]
\[ \sigma^2 = (y - X \beta)^T (y - X \beta) / n \]

\[ \beta \sim N \left( (\sigma^{-2} X^T X + V_b^{-1})^{-1} (\sigma^{-2} X^T \hat{y} + V_b^{-1} \bar{b}_0), \right. \]
\[ \left. (\sigma^{-2} X^T X + V_b^{-1})^{-1} \right) \]

\[ \sigma^2 \sim IG \left( s_1 + \frac{n}{2}, s_2 + \frac{1}{2} (\hat{y} - Xb)^T (\hat{y} - Xb) \right) \]
Assumptions of Linear Model

- Homoskedasticity
- No error in X variables
- Error in Y variables is measurement error
- Normally distributed error
- Observations are independent
- No missing data
Graph notation

- Focuses on relationships among parameters and data sets rather than distributions
- Can facilitate writing conditional distributions

\[
X \sim N(\mu, \sigma^2) \\
\mu \sim N(\mu_0, V_\mu) \\
\sigma^2 \sim IG(s_1, s_2) \\
\mu_0, V_I, S_1, S_2
\]
Linear Regression

\[ \tilde{y} \sim N(X\beta, \sigma^2) \]
Heteroskedasticity

Graphs showing two sets of data points, labeled $y^0$ and $y^1$, plotted against $x$. The data points form a linear trend in each graph.
Solutions

1) Transform the data

   1) Pro: No additional parameters

   2) Cons: No longer modeling the original data, likelihood & process model have different meaning, backtransformation non-trivial (Jensen's Inequality)

2) Model the variance

   1) Pro: working with original data and model, no tranf.

   2) Con: additional process model and parameters (and priors)
Heteroskedasticity

\[ y \sim N(\beta_1 + \beta_2 x, (\alpha_1 + \alpha_2 x)^2) \]
Example: Linear varying SD

\[ y \sim N(\beta_1 + \beta_2 x, (\alpha_1 + \alpha_2 x)^2) \]

Likelihood (R)

```r
LnL = function(theta,x,y){
  beta = theta[1:2]
  alpha = theta[3:4]  ## was sigma = theta[3]
  -sum(dnorm(y,beta[1]+beta[2]*x,
                alpha[1]+alpha[2]*x
                ,log=TRUE))
}
```

Bayes (JAGS)

```r
model{
  for(i in 1:2) { beta[i] ~ dnorm(0,0.001)}  ## priors
  for(i in 1:2) { alpha[i] ~ dlnorm(0,0.001)}  ## was prec ~ gamma(a1,a2)
  for(i in 1:n){
    prec[i] <- 1/pow(alpha[1] + alpha[2]*x[i],2)
    y[i] ~ dnorm(mu[i],prec[i])
  }
}
```
Likelihood
Bayes
Additional thoughts on modeling variance

- Need not be linear
- Can model in terms of sd, variance, or precision
- Can vary with treatments/factors or categorical variables
  - e.g. can relax the ANOVA assumptions of equal variance among treatments
- Can vary by measurement technique, sensor, etc.
Assumptions of Linear Model

- Homoskedasticity
- No error in X variables
- Error in Y variables is measurement error
- Normally distributed error
- Observations are independent
- No missing data
Errors in Variables

Regression model assumes all the error is in the Y. Often know there is non-negligible error in the measurement of X.
Errors in Variables

\[ \mu = \beta_1 + \beta_2 x \]  
Process model

\[ y \sim N(\mu, \sigma^2) \]  
Data model for y

\[ x^{(o)} \sim N(x, \tau^2) \]  
Data model for x

\[ \beta \sim N(B_0, V_B) \]  
Prior for beta

\[ \sigma^2 \sim IG(s_1, s_2) \]  
Prior for sigma

\[ \tau^2 \sim IG(t_1, t_2) \]  
Prior for tau

\[ x \sim N(X_0, V_X) \]  
Prior for X
Errors in Variables

\[ \tilde{y} \sim N(X \hat{\beta}, \sigma^2) \]
\[ x^{(o)} \sim N(x, \tau^2) \]
Full Posterior

\[ p(\vec{\beta}, \sigma^2, \tau^2, X|\bar{y}, X^{(o)}) \propto N(y|\beta_0 + \beta_1 x, \sigma^2) N(x^{(o)}|x, \tau^2) \ N(\vec{\beta}|B_0, V_B) \]

\[ IG(\sigma^2|s_1, s_2) \ IG(\tau^2|t_1, t_2) \ N(x|X_0, V_X) \]

Conditionals

\[ p(\vec{\beta}|...) \propto N(y|\beta_0 + \beta_1 x, \sigma^2) N(\vec{\beta}|B_0, V_B) \]

\[ p(\sigma^2|...) \propto N(y|\beta_0 + \beta_1 x, \sigma^2) IG(\sigma^2|s_1, s_2) \]

\[ p(\tau^2|...) \propto N(x^{(o)}|x, \tau^2) IG(\tau^2|t_1, t_2) \]

\[ p(X|...) \propto N(x^{(o)}|x, \tau^2) N(y|\beta_0 + \beta_1 x, \sigma^2) N(x|X_0, V_X) \]
Conceptually within the MCMC

- Update the regression model given the current values of X
- Update the observation error in X based on the difference between the current and observed values of X
- Update the values of X based on the observed values of X and the regression model
- Overall, integrate over the possible values of X
model {
    ## priors
    for(i in 1:2) { beta[i] ~ dnorm(0,0.001)}
    sigma ~ dgamma(0.1,0.1)
    tau ~ dgamma(0.1,0.1)
    for(i in 1:n) { x[i] ~ dunif(0,10)}

    for(i in 1:n){
        xo[i] ~ dnorm(x[i],tau)
        y[i] ~ dnorm(mu[i],sigma)
    }
}
Additional Thoughts on EIV

\[ x^{(o)} \sim g(x|\theta) \]

- Errors in X's need not be Normal
- Errors need not be additive
- Can account for known biases

\[ x^{(o)} \sim N(\alpha_0 + \alpha_1 x, \tau^2) \]
Additional Thoughts on EIV

\[ x^{(o)} \sim g(x|\theta) \]

- Errors in X's need not be Normal
- Errors need not be additive
- Can account for known biases
  \[ x^{(o)} \sim N(\alpha_0 + \alpha_1 x, \tau^2) \]
- Observed data can be a different type (proxy)
- Very useful to have informative priors
Growth Response to Moisture

- Obs
- True
- no EIV
- EIV

Growth (cm/yr)

Soil Moisture (m³/m³)
Latent Variables

- Variables that are not directly observed
- Values are inferred from model
  - Parameter model: prior on value
  - Data and Process models provide constraint

\[
p(X|\ldots) \propto N(y|\beta_0 + \beta_1 x, \sigma^2) N(x^{(o)}|x, \tau^2) N(x|X_0, V_x)
\]

- MCMC integrates over (by sampling) the values the unobserved variable could take on
- Contribute to uncertainty in parameters, model
- Ignoring this variability can lead to falsely overconfident conclusions
Assumptions of Linear Model

- Homoskedasticity
- No error in X variables
- Error in Y variables is measurement error
- Normally distributed error
- Observations are independent
- No missing data