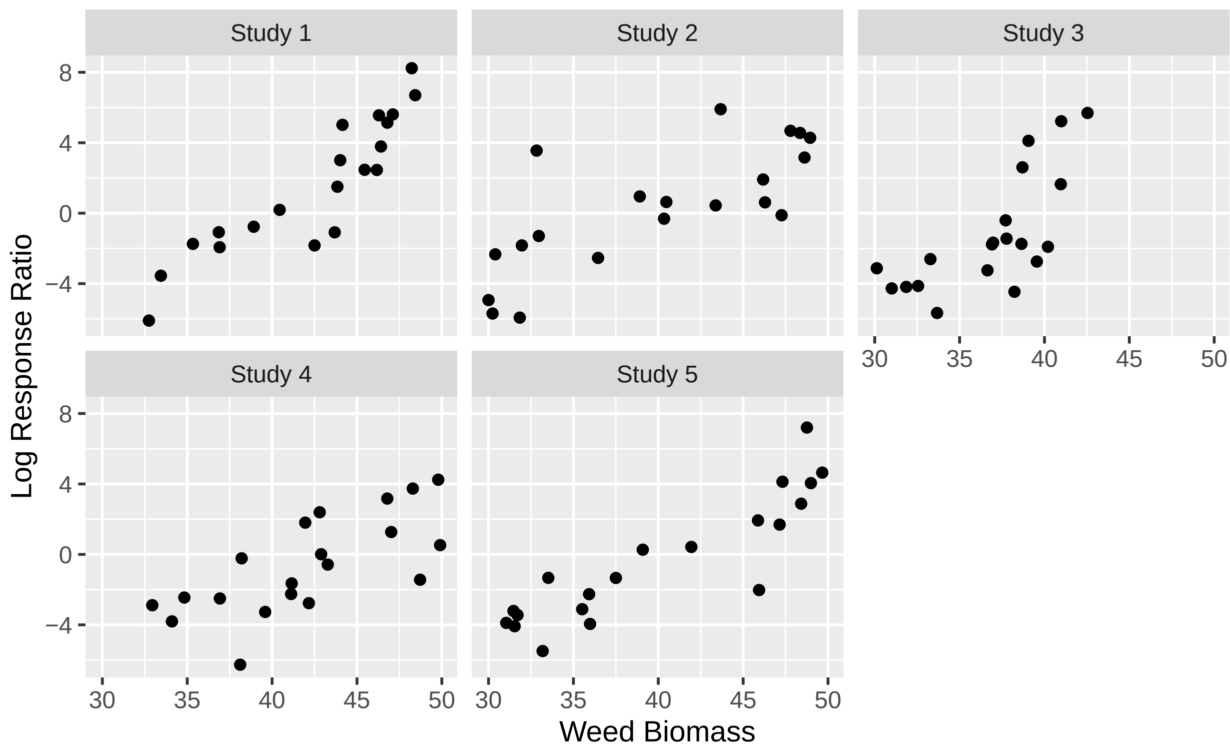
Standard Error for Inverse Prediction with Random Effect Models

AES Statistical Consulting
Consultant: Katherine Goode
Client: Gina Nichols, graduate student in agronomy (vnichols@iastate.edu)
PI: Matt Liebman (mliebman@iastate.edu)
December 27, 2019

Data, Model, and Inverse Prediction

Gina is working on a meta analysis where the response variable is the log response ratio (LRR) associated with yield. She is interested in comparing this to weed biomass. As is typical in a meta-analysis, she has data from various studies. I do not have access to the data, so I generated an example dataset to work with in this document. A plot of this data is shown below.



The model that Gina is using has the form

$$\text{LRR}_i = \beta_0 + \beta_1 \text{weed biomass}_i + \alpha_i + \epsilon_i$$

where $\alpha_i \sim N\left(0, \sigma^2_{study}\right)$ is a study random effect and $\epsilon_i \sim N\left(0, \sigma^2_{error}\right)$ is the error term for $i = 1, ..., n$. I fit this model to the example data. The code and summary of the model are included below.

```
m <- lmer(lrr ~ weed_biomass + (1|study), data = ex_data)
summary(m)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: lrr ~ weed_biomass + (1 | study)
##    Data: ex_data
```

```
##
## REML criterion at convergence: 442.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.10694 -0.52217 -0.05931  0.51334  3.14939
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  study    (Intercept) 0.5045   0.7103
##  Residual             4.4851   2.1178
## Number of obs: 100, groups:  study, 5
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  -18.63424    1.53737  -12.12
## weed_biomass   0.46152    0.03702   12.47
##
## Correlation of Fixed Effects:
##             (Intr)
## weed_biomss -0.969
```

Gina is interested in predicting the weed biomass given a LRR. In particular, she is interested in the case when LRR = -0.69, which corresponds to 50% weed control. Additionally, she would like the standard error associated with the predicted weed biomass. The estimate of the weed biomass for a given LRR can be computed using inverse prediction as

$$\hat{p}_{\text{weed control}} = \frac{\text{LRR} - \hat{\beta}_0}{\hat{\beta}_1}.$$

In order to obtain the standard error for $\hat{p}_{\text{weed control}}$, we will use the delta method.

Delta Method

Overview of the Delta Method Theory

Adapted from Dr. Mark Kaiser's Stat 520 course notes from fall 2016:

- Let $\hat{\boldsymbol{\theta}}_n$ be a sequence of asymptotically normal estimators of a parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$ with mean $\boldsymbol{\theta}$ and covariance matrix $c_n^2\boldsymbol{\Sigma}$.
- Let $g(\boldsymbol{\theta})$ be a real-valued function of $\boldsymbol{\theta}$ that is continuously differentiable in a neighborhood of $\boldsymbol{\theta}$.
- Let $\mathbf{d}$ be a vector of length p with a jth element of

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_j}.$$

- Then $g\left(\hat{\boldsymbol{\theta}}\right)$ is asymptotically normal with mean $g(\boldsymbol{\theta})$ and covariance matrix $c_n^2\mathbf{d}\boldsymbol{\Sigma}\mathbf{d}'$.
- In likelihood estimation and inference, $\hat{\boldsymbol{\theta}}_n$ can be used as a plug-in estimator of $\boldsymbol{\theta}$ to estimate $g(\boldsymbol{\theta})$ and $c_n^2\mathbf{d}\boldsymbol{\Sigma}\mathbf{d}'$.
- Additionally, in the iid case, $c_n^2\boldsymbol{\Sigma}$ is the inverse total information matrix. That is

$$c_n^2 \boldsymbol{\Sigma} = \frac{1}{n} I(\boldsymbol{\theta})^{-1}.$$

This can be estimated by the negative inverse Hessian matrix, which is the estimated variance-covariance matrix of $\boldsymbol{\theta}$.

- Thus, the variance of $g\left(\hat{\boldsymbol{\theta}}\right)$ can be obtained as

$$Var\left(g\left(\hat{\boldsymbol{\theta}}\right)\right) = \mathbf{d} Cov\left(\hat{\boldsymbol{\theta}}\right) \mathbf{d}',$$

and the standard error of $g\left(\hat{\boldsymbol{\theta}}\right)$ can be computed as

$$SE\left(g\left(\hat{\boldsymbol{\theta}}\right)\right) = \sqrt{Var\left(g\left(\hat{\boldsymbol{\theta}}\right)\right)}.$$

A good reference on the delta method that also describes how to apply the delta method in R can be found on the IDRE website.

Notation and Derivatives for the Weed Biomass Problem

In our case, $\boldsymbol{\theta} = (\beta_0, \beta_1)$ and

$$g(\boldsymbol{\theta}) = \frac{\text{LRR} - \beta_0}{\beta_1}.$$

To obtain $\mathbf{d}$, we need to compute the partial derivatives of $g(\boldsymbol{\theta})$ in terms of β_0 and β_1. These can be computed as follows.

$$\begin{aligned}
\frac{\partial g(\boldsymbol{\theta})}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0}\left(\frac{LRR - \beta_0}{\beta_1}\right) \\
&= \frac{\partial}{\partial \beta_0}\left(\frac{LRR}{\beta_1}\right) - \frac{\partial}{\partial \beta_0}\left(\frac{\beta_0}{\beta_1}\right) \\
&= 0 - \frac{1}{\beta_1} \\
&= -\frac{1}{\beta_1}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial g(\boldsymbol{\theta})}{\partial \beta_1} &= \frac{\partial}{\partial \beta_1}\left(\frac{LRR - \beta_0}{\beta_1}\right) \\
&= \frac{\partial}{\partial \beta_1}\left(LRR - \beta_0\right)\left(\beta_1^{-1}\right) \\
&= -\left(LRR - \beta_0\right)\left(\beta_1^{-2}\right) \\
&= \frac{-LRR + \beta_0}{\beta_1^2}
\end{aligned}$$

Thus,

$$\mathbf{d} = \begin{bmatrix} -\frac{1}{\beta_1} & \frac{-LRR+\beta_0}{\beta_1^2} \end{bmatrix} \tag{1}$$

The standard error for $g(\hat{\boldsymbol{\theta}})$ will be

$$\sqrt{Var[g(\hat{\boldsymbol{\theta}})]} = \left(\mathbf{d}Cov\left[\hat{\boldsymbol{\theta}}\right]\mathbf{d}'\right)^{1/2}$$

where

$$Cov\left[\hat{\boldsymbol{\theta}}\right] = \begin{bmatrix} Var\left[\hat{\beta}_0\right] & Cov\left[\hat{\beta}_0, \hat{\beta}_1\right] \\ Cov\left[\hat{\beta}_0, \hat{\beta}_1\right] & Var\left[\hat{\beta}_1\right] \end{bmatrix}.$$

Applying the Delta Method to the Inverse Prediction

R Function

I wrote the function `compute_se` to implement the delta method computations to compute the standard error for the weed biomass prediction for a given LRR. The function also returns the estimate of the LRR and a 95% confidence interval for the prediction. The inputs and outputs of the function are as follows.

Inputs:

- `lrr`: LRR for which to compute the weed biomass
- `betas`: estimated regression coefficients of β_0 and β_1 (should be a vector of length 2)
- `vcov`: estimated variance covariance matrix of β_0 and β_1 (should be a 2x2 matrix)

Outputs:

- data frame with the variables of
  - `lrr`: age that was specified for the computations
  - `pred_biomass`: estimated weed biomass for the specified `lrr`
  - `se`: standard error for the estimated weed biomass (computed using the delta method)
  - `ci_Lower`: lower bound of the 95% confidence interval for weed biomass
  - `ci_Upper`: upper bound of the 95% confidence interval for weed biomass

The code for the function `compute_se` is included below.

```r
# Function for computing the delta-method standard error of weed biomass
compute_se <- function(lrr, betas, vcov){

  # Separate the betas
  b0 <- betas[1]
  b1 <- betas[2]

  # Compute the inverse prediction of weed biomass
  pred_biomass <- (lrr - b0) / b1

  # Create an empty 1x2 matrix to store the elements of d in
  d <- matrix(NA, nrow = 1, ncol = 2)

  # Compute the elements of d (partial derivatives of g(beta))
  d[1] <- -1 / b1
  d[2] <- (-lrr + b0) / (b1^2)

  # Compute the standard error of annual survival (using the delta method)
  se <- sqrt(d %*% vcov %*% t(d))

  # Compute the lower and upper bounds of the 95% CI for annual survival
```

```r
  lower <- pred_biomass - (1.96 * se)
  upper <- pred_biomass + (1.96 * se)

  # Return the log response ratio, the predicted weed biomass given the
  # resopnse ration, the delta method standard error, and the lower
  # and upper bounds of the 95% CI for the predicted weed biomass
  return(data.frame(lrr = lrr,
                    pred_biomass = pred_biomass,
                    se = se,
                    ci_lower = lower,
                    ci_upper = upper))

}
```

Standard Error Calculation for LRR = -0.69

I first extracted the vector of estimated betas and the variance-covariance matrix.

```r
# Extract the variance-covariance matrix for beta0 and beta1
betas <- as.vector(summary(m)$coefficients[,1])
betas
```

```
## [1] -18.6342374   0.4615191
```

```r
# Extract the variance-covariance matrix for beta0 and beta1
vcov <- matrix(vcov(m), nrow = 2)
vcov
```

```
##             [,1]         [,2]
## [1,]  2.36351420 -0.055124717
## [2,] -0.05512472  0.001370177
```

Then I applied the `compute_se` function to compute the weed biomass estimate, standard error, and 95% confidence interval for a LRR of -0.69. The results are shown below.

```r
# Apply the compute_se function
res <- compute_se(lrr = -0.69, betas = betas, vcov = vcov)
res
```

```
##     lrr pred_biomass        se ci_lower ci_upper
## 1 -0.69     38.88081 0.8342623 37.24566 40.51596
```