Bayesian Linear Regression

Bayesian Data Analysis

Steve Buyske

Frequentist Framework for the Linear Model

- *Linear model* is another term for linear regression.
- In the frequentist framework, we might write

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad \text{where } \epsilon_i \sim \text{Normal}(0, \sigma) \text{ and independent.}$$

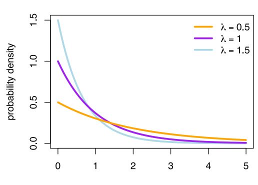
- We might instead write the mean-centered version,

$$y_i = \alpha' + \beta(x_i - \bar{x}) + \epsilon_i, \qquad \text{where } \epsilon_i \sim \text{Normal}(0, \sigma) \text{ and independent.}$$

  - This gives a more natural interpretation of the intercept when $x_i = 0$ doesn't make much sense (like a regression of weight on height).
  - In the first version, when $x_i = 0$, the predicted value if $\hat{y}_i = \alpha$.
  - In the first version, when $x_i = \bar{x}$, the predicted value if $\hat{y}_i = \alpha'$.
  - (You might remember in this case that $\hat{\alpha}' = \bar{y}$.)

Bayesian Framework for the Linear Model

- Model:
  - $y_i \sim \text{Normal}(\mu_i, \sigma)$
  - $\mu_i = \alpha + \beta x_i$ *this is sometimes called the link*
  - together, *these will give us the likelihood*.
- The priors:
  - $\alpha \sim \text{Normal}(\text{something}, \text{something})$,
  - $\beta \sim \text{Normal}(0, \text{something})$,
  - $\sigma \sim \text{Exponential}(\text{something})$.
  - (These are not the only possible priors, just a common set.)

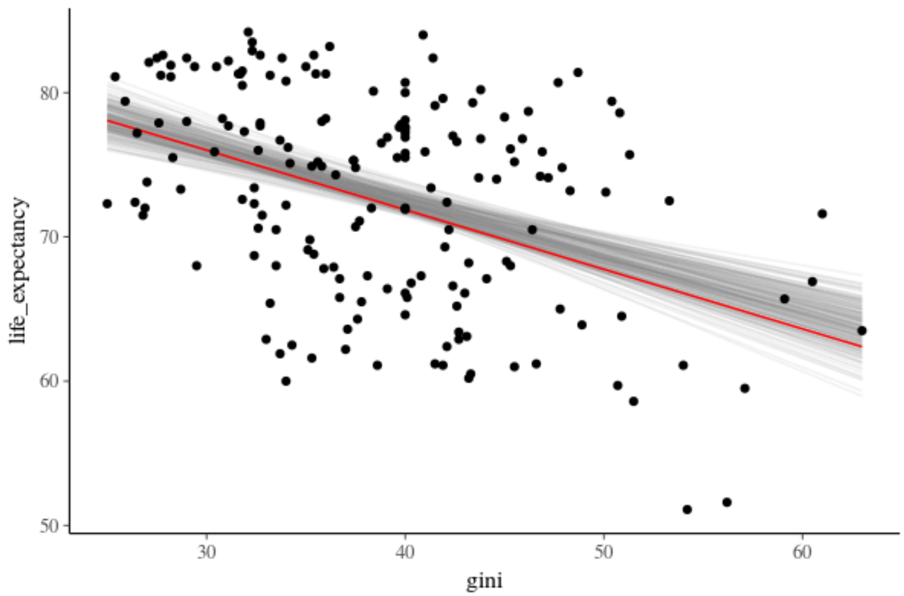


Priors on the gini by life expectancy example

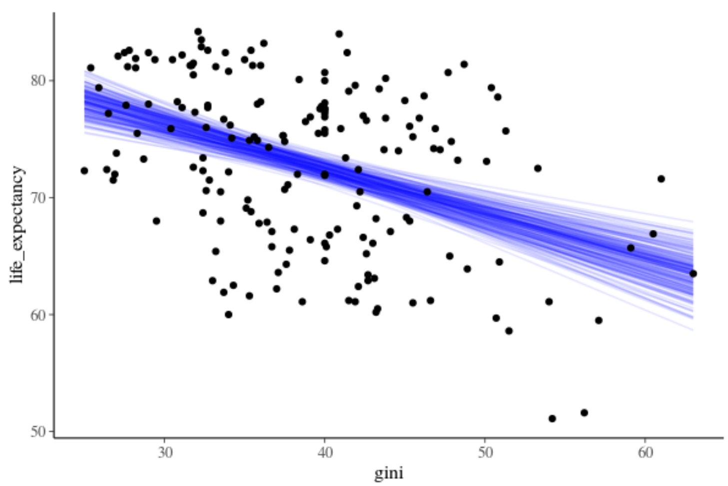
- The priors:
    - $\alpha \sim \text{Normal}(73, 18)$,
    - $\beta \sim \text{Normal}(0, 2.3)$,
    - $\sigma \sim \text{Exponential}(0.14)$.
- These priors are known as *weakly informative*—they are pretty dispersed, but they do have some information in them.
- Presumably you're quite familiar with the normal distribution, but you can visualize various distributions at https://ben18785.shinyapps.io/distribution-zoo/

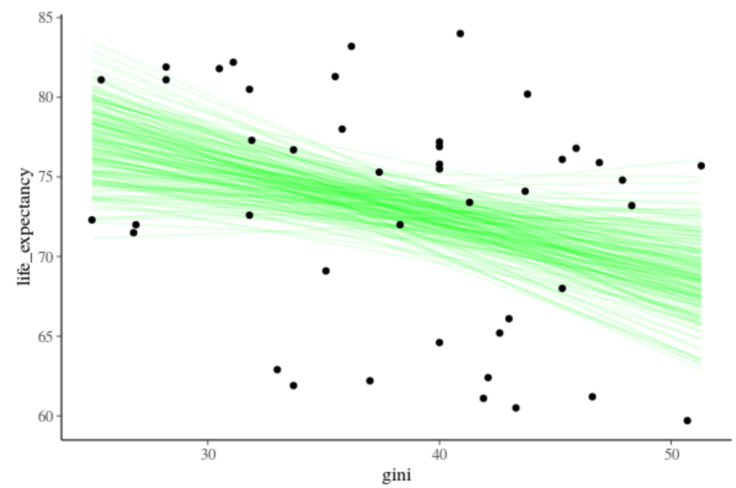
The Parameter is a Parameter Vector

- Keep in mind that we have a vector of parameters: (α, β, σ).
  - This means our posterior is actually 3-dimensional.
- A single draw from the posterior corresponds to
  - a regression line, because of the (α, β), and
  - the spread of noise around that line, because of σ.
- The plot shows lines corresponding to draws from the posterior.



- The purple plot on the left shows lines corresponding to draws from the posterior.
- The green plot on the right shows lines from the posterior if we reduce the dataset to a smaller number of observations (notice there are fewer dots).
    - You would never do that in practice—it's just to illustrate a point.
- You can see that there is more variability in the lines in the green plot, with fewer observations, than the purple plot.
- This shows how the more evidence (i.e., the more data), the tighter the posterior will be.





Model fit

- There are two separate model fitting questions in a Bayesian data analysis.
- 1. Did the MCMC process work? That is, does the MCMC sample do a good job approximating the actual posterior distribution?
  - We will use the chains to help us decide.
  - If we think there is a problem,
    - our main recourse is to run the chains for longer,
    - or (sometimes) to change the proposal distribution by changing the step size.
  - We will look at this is detail in the future.

Model fit cont.

- 2. Does the model fit the data?
  - In the frequentist framework, we usually start with residual plots, where for a particular x_i, we define the residual as $e_i = \hat{y}_i - y_i$.
  - In the Bayesian framework, for a particular x_i, there is actually of distribution for the residual e_i, since $\hat{y}_i$ itself has a distribution.
    - Why? Because $\hat{y}_i$ depends on α and β, each of which has a distribution.
- It is often simplest just to consider the residuals from a frequentist analysis.
- We will look at Bayesian residual analysis in the future.

Posterior Predictive Check

- A diagnostic unique to the Bayesian world is the *posterior predictive check*.
- The posterior predictive distribution is the distribution of the outcome variable given the predictors and the posterior of the parameters.
- The idea is to make a draw from the posterior of the parameters and then, based on the predictor variables, simulate the values of the outcome.
- Do that repeatedly.
- Compare the actual distribution of the outcome variable (in black in the plot below) with the many simulated distributions (in blue) just created.
- We will mostly plot the distributions directly, but one could also look at the distribution of the mean, say.