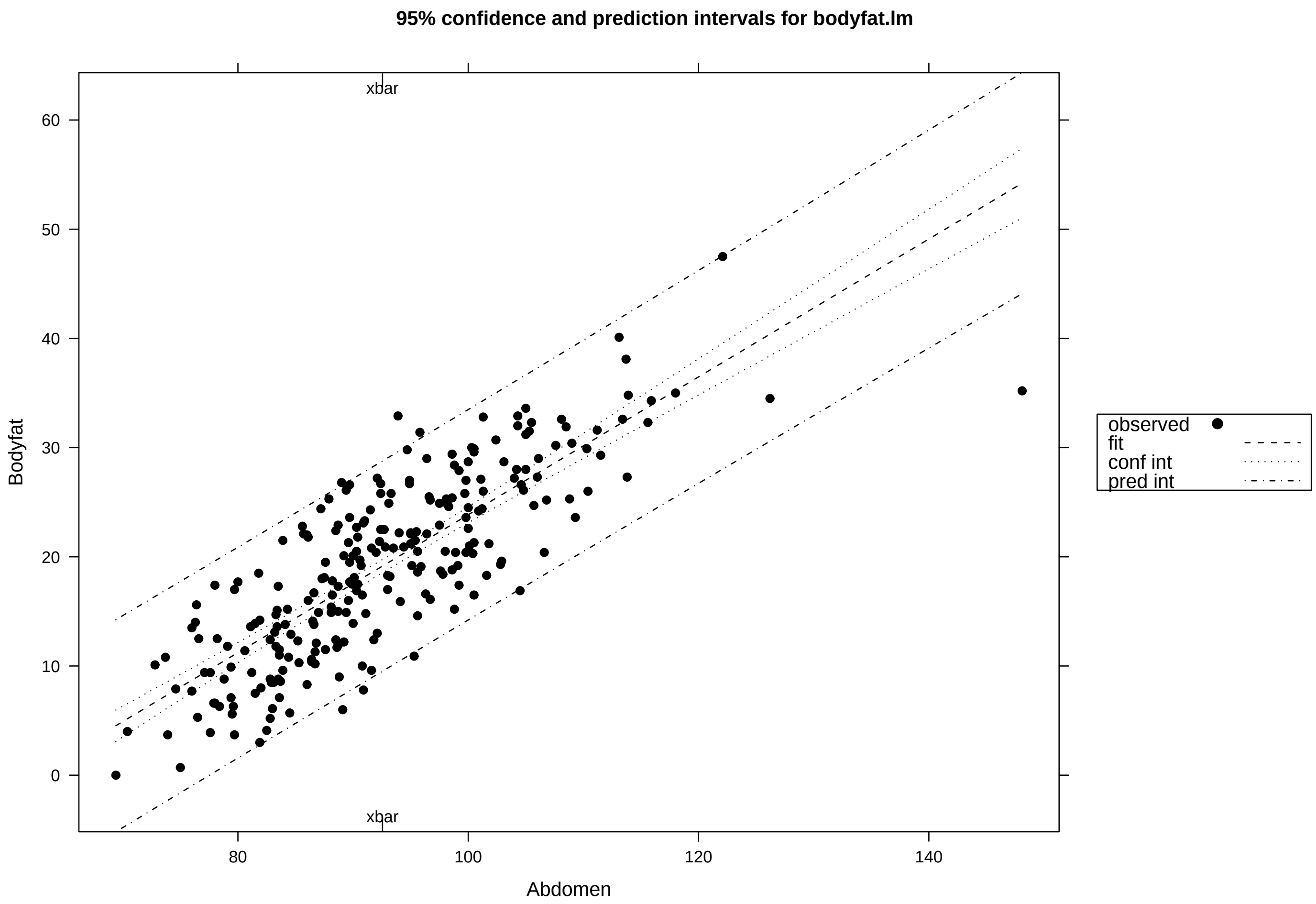
Robust Bayesian Simple Linear Regression

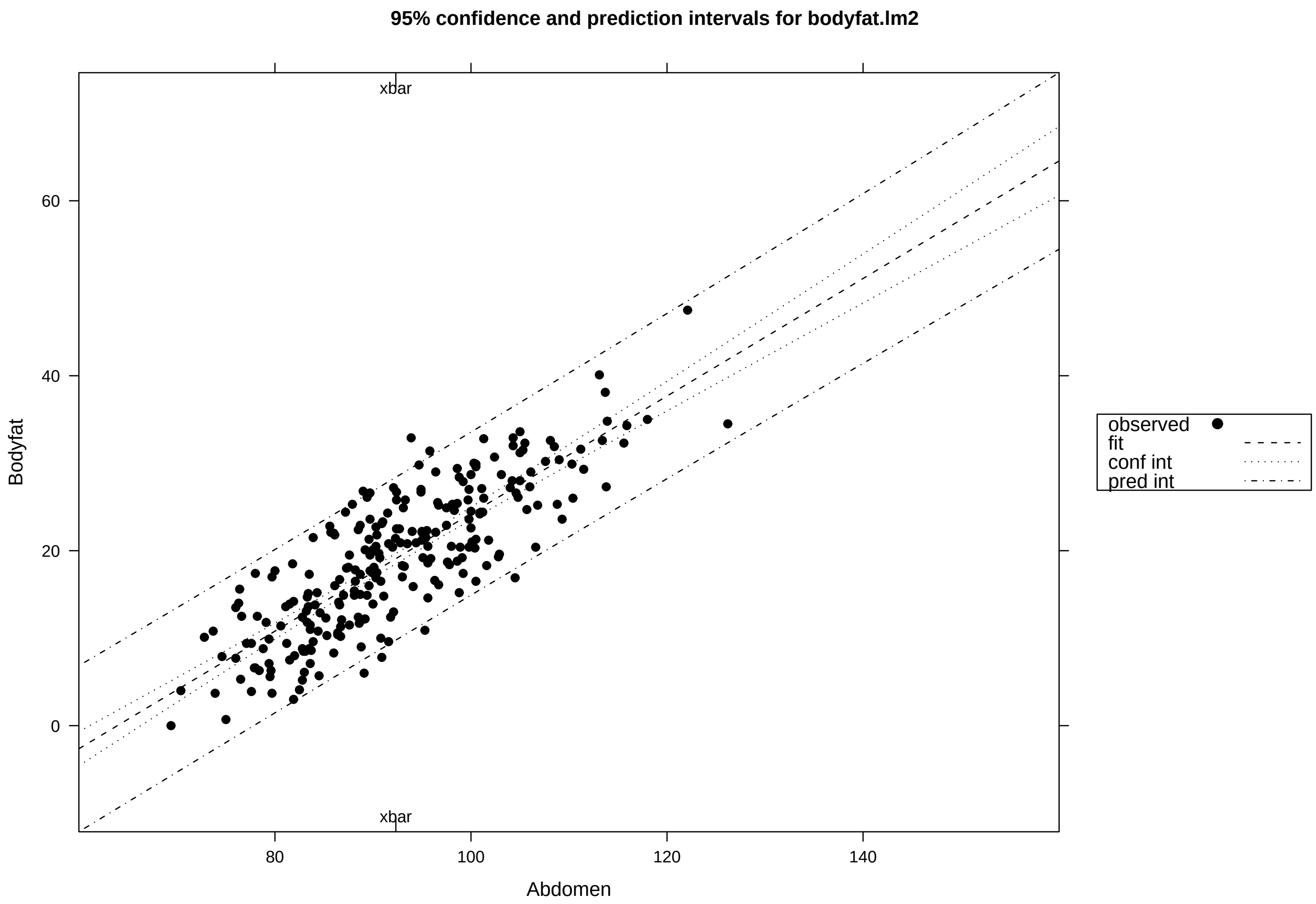
October 1, 2008

Readings: GIll 4

Body Fat Data: Intervals w/ All Data



Intervals: without case 39



Interpretations

- For a given Abdominal circumference, our probability that the mean bodyfat percentage is in the intervals given by the dotted lines is 0.95.
- For a new man with a given Abdominal circumference, our probability that his bodyfat percentage is in the intervals given by the dashed lines is 0.95.
- Both have same point estimate
- Increased uncertainty for prediction of a new observation versus estimating the expected value.

Which analysis do we use? with Case 39 or not – or something different?

Options for Handling Influential Cases

- Are there scientific grounds for eliminating the case?
- Test if the case has a different mean than population
- Report results with and without the case
- Determine if transformations (Y and/or X) reduce influence
- Add other predictors
- Change error assumptions:

$$\varepsilon_i \overset{iid}{\sim} t(\nu, 0, 1)\,\sigma$$

Robust Regression using heavy tailed error distributions

Likelihood & Posterior

$$Y_i \overset{ind}{\sim} t(\nu, \alpha + \beta x_i, 1/\phi)$$

$$L(\alpha, \beta, \phi) \propto \prod_{i=1}^{n} \phi^{1/2} \left(1 + \frac{\phi(y_i - \alpha - \beta x_i)^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

Use Prior $p(\alpha, \beta, \phi) \propto 1/\phi$

Posterior distribution

$$p(\alpha, \beta, \phi \mid Y) \propto \phi^{n/2-1} \prod_{i=1}^{n} \left(1 + \frac{\phi(y_i - \alpha - \beta x_i)^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

No closed formed expressions!

Scale-Mixtures of Normal Representation

$$Z_i \stackrel{iid}{\sim} t(\nu, 0, \sigma^2) \Rightarrow$$

$$Z_i \mid \lambda_i \stackrel{ind}{\sim} N(0, \sigma^2/\lambda_i)$$

$$\lambda_i \stackrel{iid}{\sim} G(\nu/2, \nu/2)$$

Integrate out "latent" λ's to obtain marginal distribution.

Latent Variable Model

$$Y_i \mid \alpha, \beta, \phi, \lambda \overset{ind}{\sim} N(\alpha + \beta x_i, \frac{1}{\phi \lambda_i})$$

$$\lambda_i \overset{iid}{\sim} G(\nu/2, \nu/2)$$

$$p(\alpha, \beta, \phi) \propto 1/\phi$$

Joint Posterior Distribution:

$$p((\alpha, \beta, \phi, \lambda_1, \ldots, \lambda_n \mid Y) \propto \phi^{n/2-1} \exp\left\{ -\frac{\phi}{2} \sum \lambda_i (y_i - \alpha - \beta x_i)^2 \right\} \times \prod_{i=1}^{n} \lambda_i^{\nu/2-1} \exp(-\lambda_i \nu/2)$$

Posterior Distributions

First Factorization:

- $\alpha, \beta, \phi \mid \lambda_1, \ldots, \lambda_n$ has a Normal-Gamma distribution
- Marginal distribution of $\lambda_1, \ldots, \lambda_n$ (given the data) is hard!

Second Factorization:

- $\lambda_1, \ldots, \lambda_n \mid \alpha, \beta, \phi$ independent Gamma
- Marginal Distribution α, β, ϕ given the data is hard!

Can we combine the easy (posterior) distributions ???

$$\alpha, \beta, \phi \mid \lambda_1, \ldots, \lambda_n, Y$$

$$\lambda_1, \ldots \lambda_n \mid \alpha, \beta, \phi, Y$$

Answer is a Qualified

Yes!
While the product of the two conditional distributions is not equal to the joint posterior distribution (unless they are independent), we can create a scheme to sample from the two distributions that ensures that after a sufficient number of samples that the subsequent samples represent a (dependent) sequence of draws from the joint posterior distribution!

The easiest version is the single component Gibbs Sampler.

Single Component Gibbs Sampler

Start with $(\alpha^{(0)}, \beta^{(0)}, \phi^{(0)}, \lambda_1^{(0)}, \ldots, \lambda_n^{(0)})$
For $t = 1, \ldots, T$, generate from the following sequence of Full Conditional distributions:

- $p(\alpha \mid \beta^{(t-1)}, \phi^{(t-1)}, \lambda_1^{(t-1)}, \ldots, \lambda_n^{(t-1)}, Y)$
- $p(\beta \mid \alpha^{(t)}, \phi^{(t-1)}, \lambda_1^{(t-1)}, \ldots, \lambda_n^{(t-1)}, Y)$
- $p(\phi \mid \alpha^{(t)}, \beta^{(t)}, \phi^{(t-1)}, \lambda_1^{(t-1)}, \ldots, \lambda_n^{(t-1)}, Y)$
- $p(\lambda_j \mid \alpha^{(t)}, \beta^{(t)}, \phi^{(t)}, \lambda_{(-j)}^{(t-1)}, Y)$ for $j = 1, \ldots, n$

$\lambda_{(-j)}$ is the vector of λs excluding the jth component

Easy to find and sample!