Cost effective prediction of bodyfat

An example of project presentation slides

Aki Vehtari
Aalto University

Cost effective prediction of bodyfat

An example of project presentation slides

Aki Vehtari
Aalto University

Introduce yourself

Measuring bodyfat percentage

- Bodyfat percentage is related to many health outcomes

[Nice figures here]

Measuring bodyfat percentage

- Bodyfat percentage is related to many health outcomes
- Relatively accurate way to measure bodyfat is to weight a person in air and immersed in water
  - proportion of body fat can be derived from body density with Siri's (1956) formula
  - water immersion requires a big tub for the water and harness system for lowering a person to water

[Nice figures here]

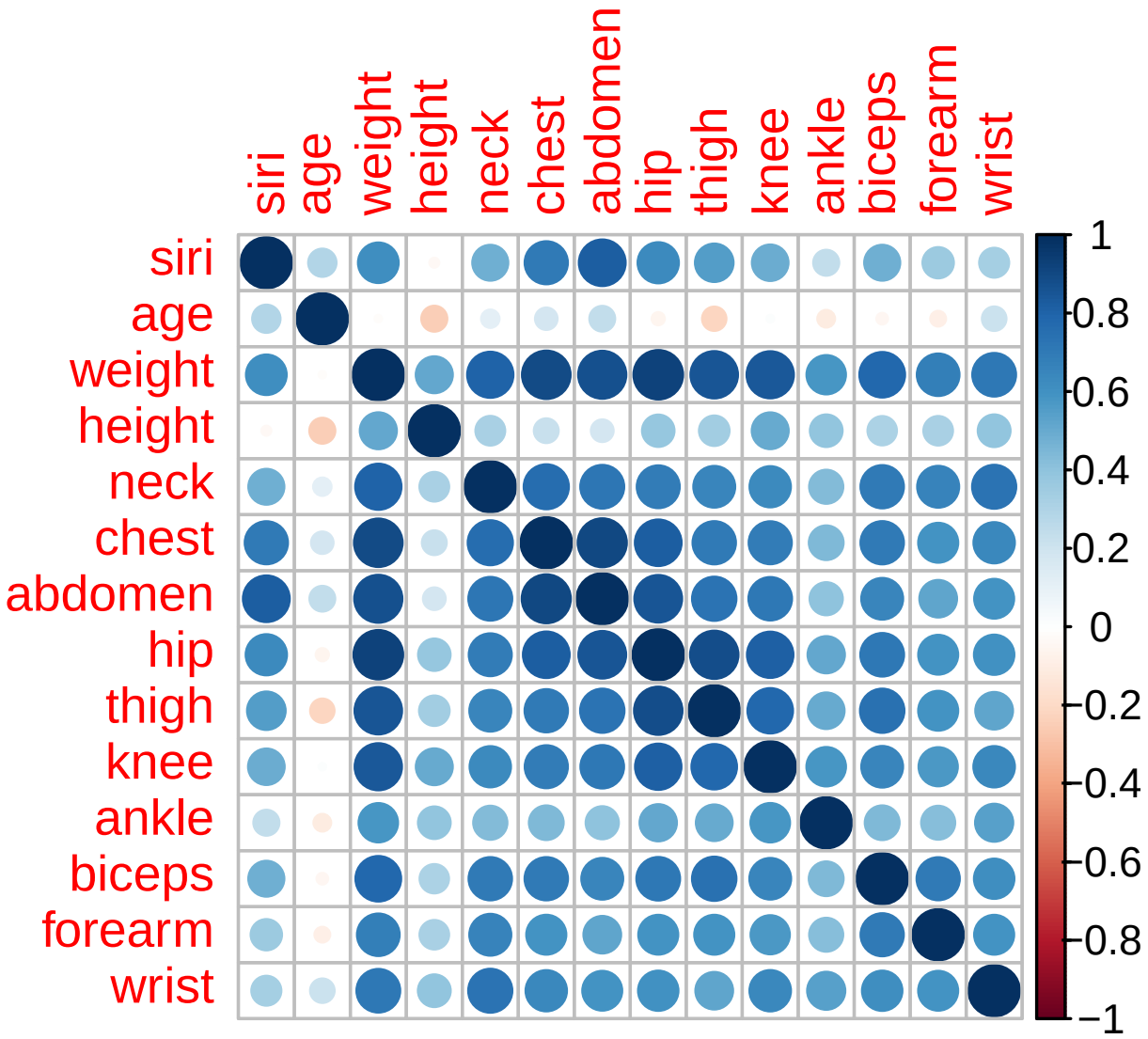
Measuring bodyfat percentage

- Bodyfat percentage is related to many health outcomes
- Relatively accurate way to measure bodyfat is to weight a person in air and immersed in water
  - proportion of body fat can be derived from body density with Siri's (1956) formula
  - water immersion requires a big tub for the water and harness system for lowering a person to water
- Can we estimate the bodyfat percentage with faster and a smaller equipment?
  - with just a scale and measure tape?

[Nice figures here]

Measuring bodyfat percentage

- With just a scale and measure tape?



Bodyfat predictive model

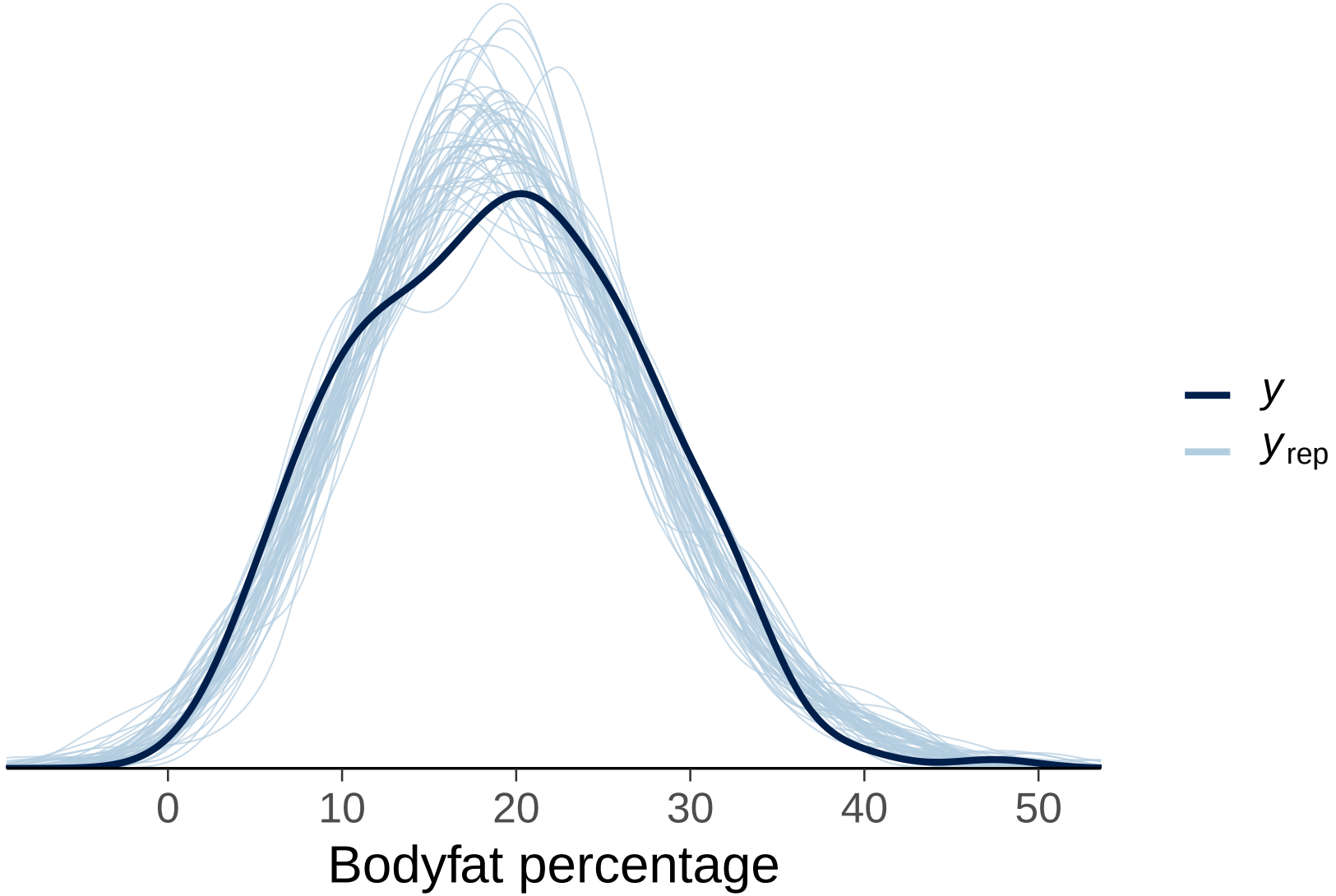
- Gaussian linear regression model with regularized horseshoe prior ($p_0 = 5$) on coefficients

Bodyfat predictive model

- Gaussian linear regression model with regularized horseshoe prior ($p_0 = 5$) on coefficients
- Model build with `rstanarm` and inference run with Stan
  - all convergence diagnostics were good

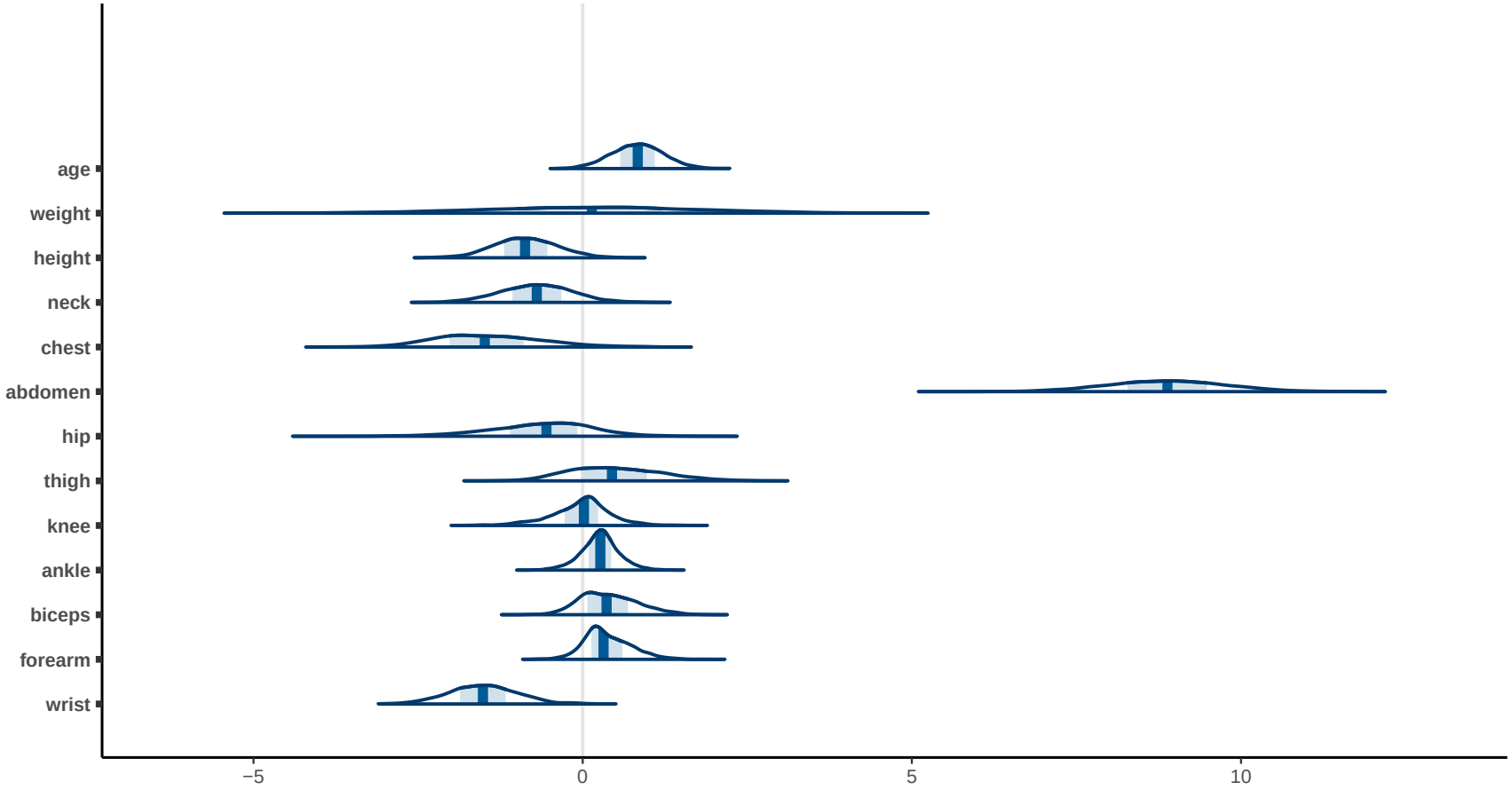
Bodyfat model checking

Posterior predictive checking



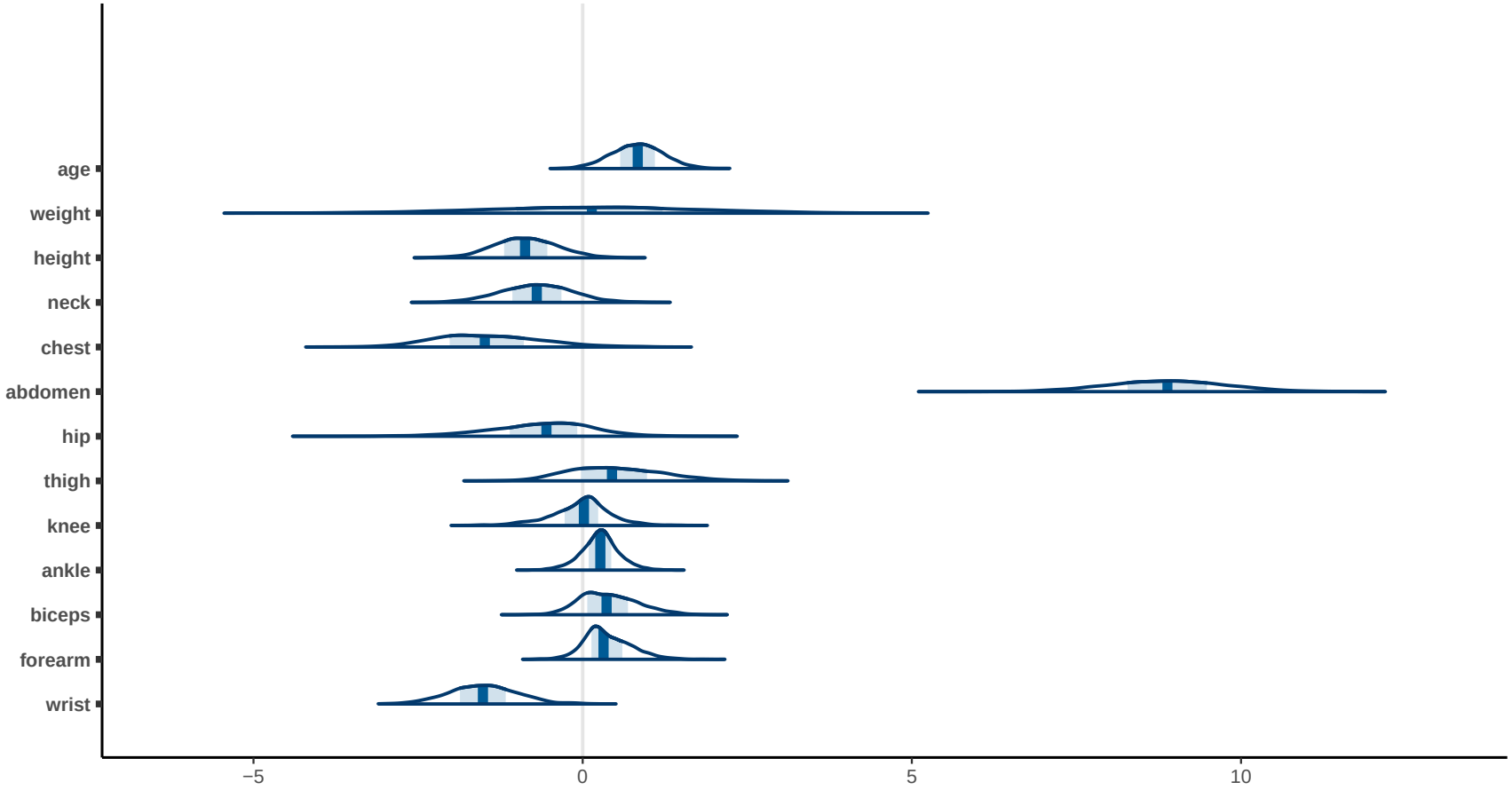
Bodyfat

Marginal posteriors of coefficients



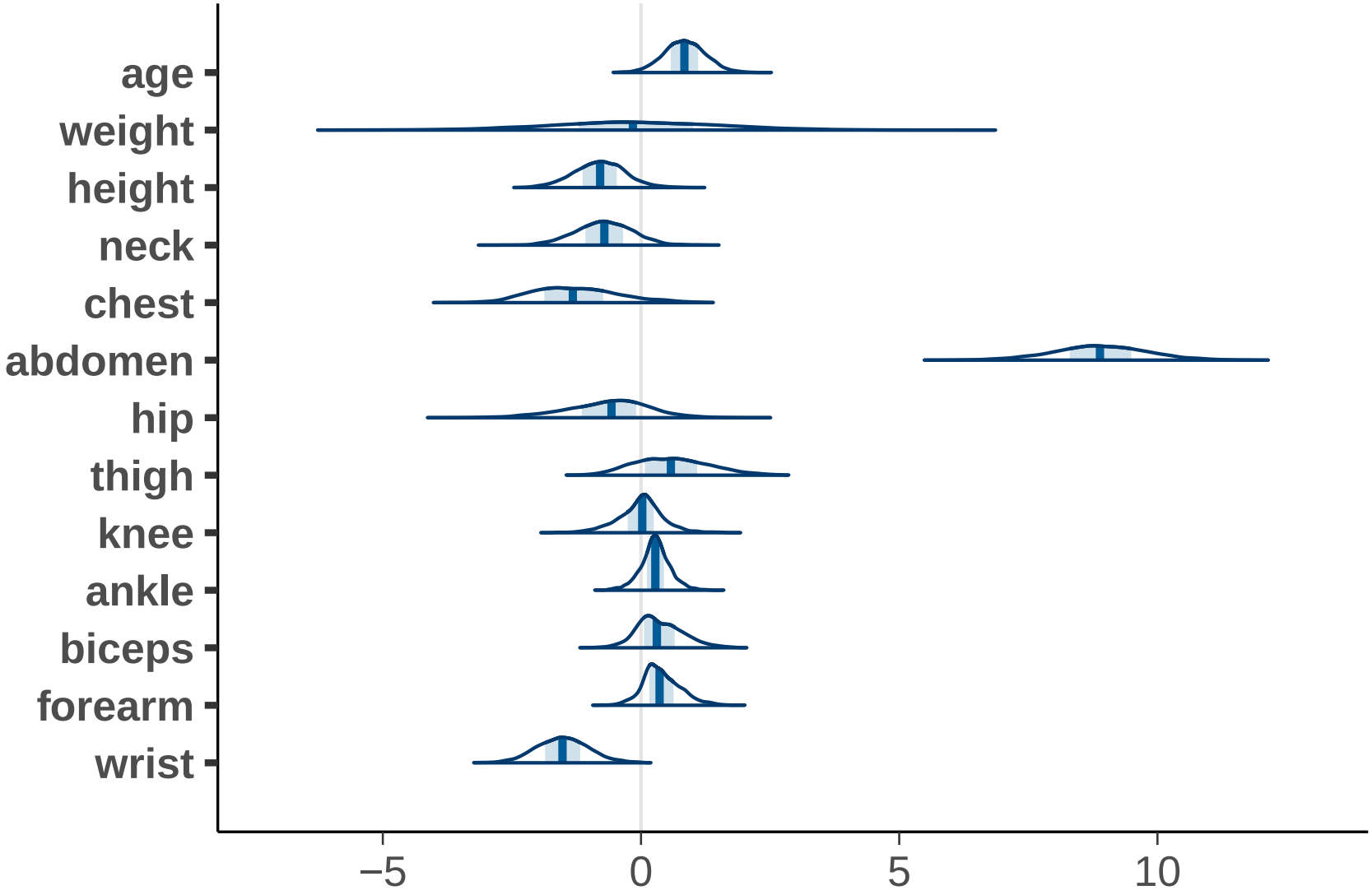
Bodyfat

Check that the font in all figures is big enough!



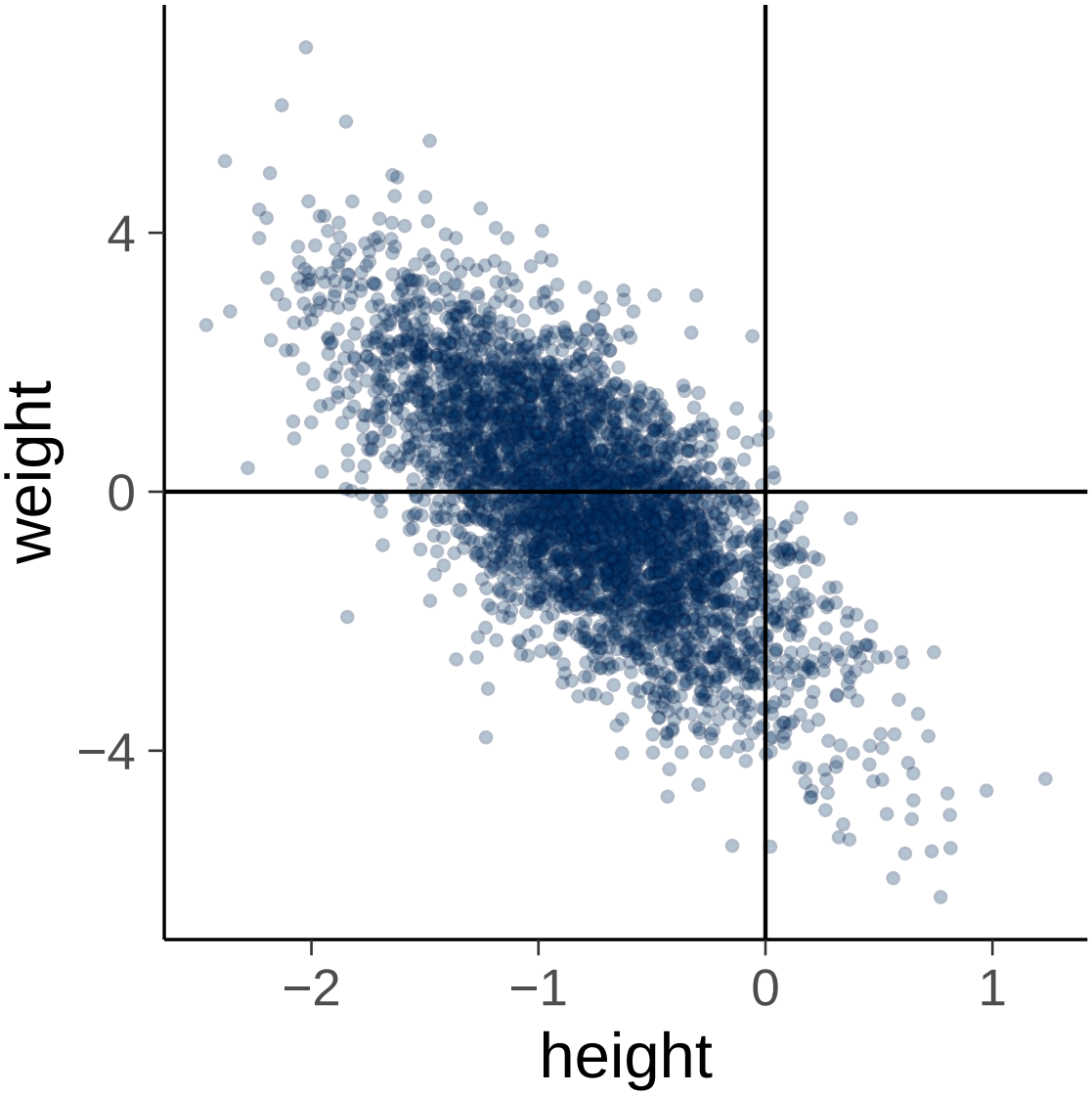
Bodyfat

Marginal posteriors of coefficients



Bodyfat

Bivariate marginal of weight and height



Bodyfat variable selection

- Do we need all the measurements?
- We find the model with a minimal set of variables which have similar predictive performance as the model with all variables

Bodyfat variable selection

- Do we need all the measurements?
- We find the model with a minimal set of variables which have similar predictive performance as the model with all variables
- We use projection predictive variable selection implemented in `projpred` package

Projective predictive covariate selection

- The full model predictive distribution represents our best knowledge about future $\tilde{y}$

$$p(\tilde{y}|D) = \int p(\tilde{y}|\theta)p(\theta|D)d\theta,$$

  where $\theta = (\beta, \sigma^2)$) and β is in general non-sparse (all $\beta_j \neq 0$)
- What is the best distribution $q_\perp(\theta)$ given a constraint that only selected covariates have nonzero coefficient
- Optimization problem:

$$q_\perp = \arg\min_q \frac{1}{n}\sum_{i=1}^{n} \mathrm{KL}\left(p(\tilde{y}_i \mid D) \,\|\, \int p(\tilde{y}_i \mid \theta)q(\theta)d\theta\right)$$

- Optimal projection from the full posterior to a sparse posterior (with minimal predictive loss)

For 10min presentation, too much information

- The full model predictive distribution represents our best knowledge about future $\tilde{y}$

$$p(\tilde{y}|D) = \int p(\tilde{y}|\theta)p(\theta|D)d\theta,$$

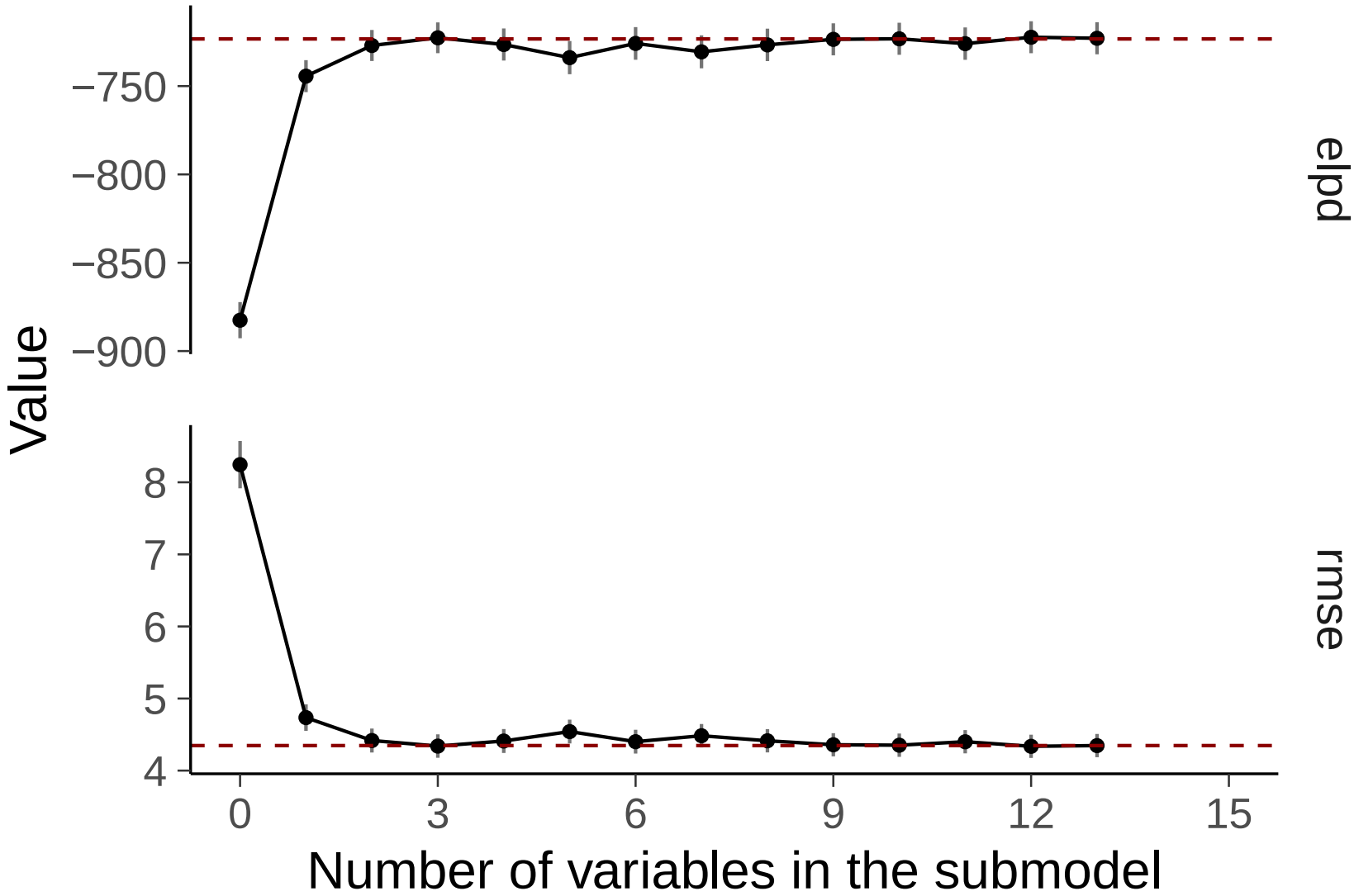
  where $\theta = (\beta, \sigma^2)$) and β is in general non-sparse (all $\beta_j \neq 0$)
- What is the best distribution $q_\perp(\theta)$ given a constraint that only selected covariates have nonzero coefficient
- Optimization problem:

$$q_\perp = \arg\min_q \frac{1}{n}\sum_{i=1}^{n} \mathrm{KL}\left(p(\tilde{y}_i \mid D) \,\|\, \int p(\tilde{y}_i \mid \theta)q(\theta)d\theta\right)$$

- Optimal projection from the full posterior to a sparse posterior (with minimal predictive loss)

Bodyfat

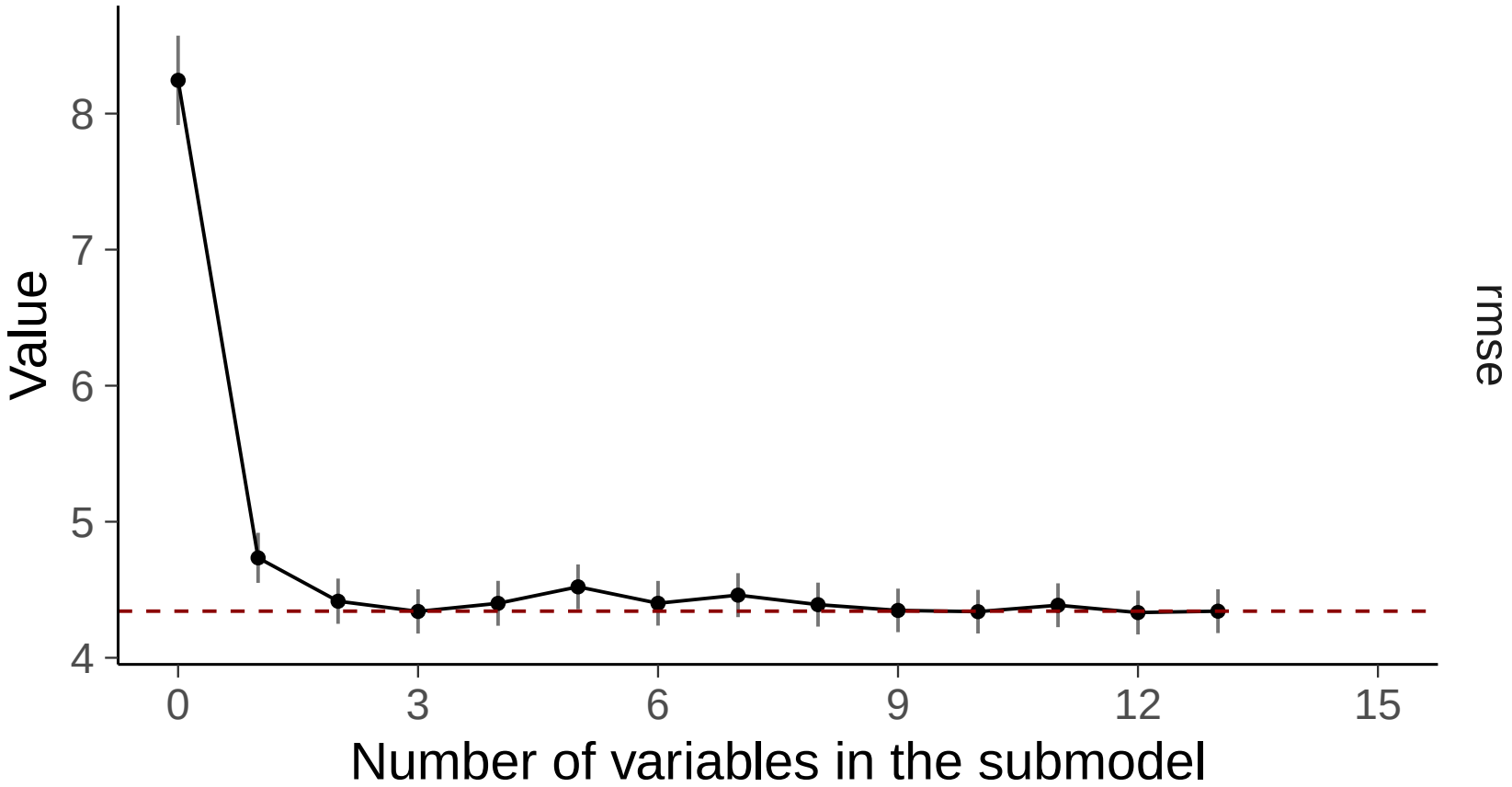
The predictive performance of the full and submodels



Bodyfat

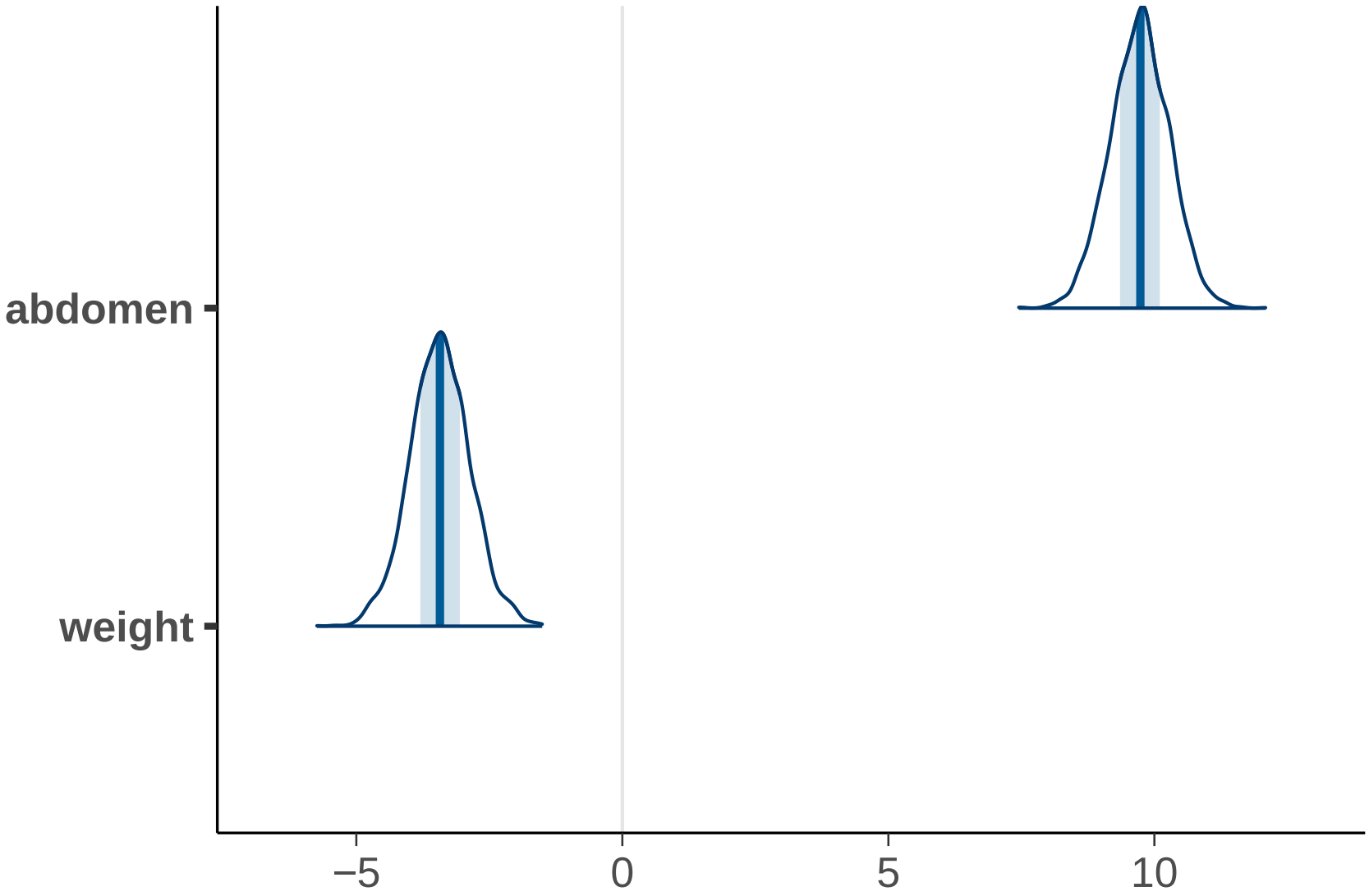
The predictive performance of the full and submodels

One of these plots is probably sufficient



Bodyfat

Marginals of projected posterior



Bodyfat – Conclusion

- Bodyfat percentage estimated using water immersion can be predicted using scale and tape measure

Bodyfat – Conclusion

- Bodyfat percentage estimated using water immersion can be predicted using scale and tape measure
- The accuracy using mean of data is 16%-units (95% interval)

Bodyfat – Conclusion

- Bodyfat percentage estimated using water immersion can be predicted using scale and tape measure
- The accuracy using mean of data is 16%-units (95% interval)
- The accuracy using all anthropometric measures is 8.6%-units (95% interval)

Bodyfat – Conclusion

- Bodyfat percentage estimated using water immersion can be predicted using scale and tape measure
- The accuracy using mean of data is 16%-units (95% interval)
- The accuracy using all anthropometric measures is 8.6%-units (95% interval)
- The same accuracy can be obtained using just abdomen circumference and weight

Bodyfat – Conclusion

- Bodyfat percentage estimated using water immersion can be predicted using scale and tape measure
- The accuracy using mean of data is 16%-units (95% interval)
- The accuracy using all anthropometric measures is 8.6%-units (95% interval)
- The same accuracy can be obtained using just abdomen circumference and weight
- More results at avehtari.github.io/modelselection/bodyfat.html

THANKS!

NO “THANKS”!

NO “THANKS”!

- Don’t ever end with a slide having just “THANKS”

NO “THANKS”!

- Don’t ever end with a slide having just “THANKS”
- “THANKS” slide has zero information content

NO “THANKS”!

- Don’t ever end with a slide having just “THANKS”
- “THANKS” slide has zero information content
- Leave the conclusion slide or contact information slide

Conclusion

- Bodyfat percentage estimated using water immersion can be predicted using scale and tape measure
- The accuracy using mean of data is 16%-units (95% interval)
- The accuracy using all anthropometric measures is 8.6%-units (95% interval)
- The same accuracy can be obtained using just abdomen circumference and weight
- More results at avehtari.github.io/modelselection/bodyfat.html

Additional information

- You can have additional slides after the conclusion for supporting material to answer questions
  - for example, in this course, include Stan code and additional convergence and model checking results

Gaussian linear model with regularized horseshoe prior

```
p0 <- 5 # prior guess for the number of relevant variables
tau0 <- p0/(p-p0) * 1/sqrt(n)
rhs_prior <- hs(global_scale=tau0)
fitrhs <- stan_glm(formula, data = df, prior=rhs_prior,
                   QR=TRUE, seed=SEED, refresh=0)
```

Additional information

- You can have additional slides after the conclusion for supporting material to answer questions
  - for example, in this course, include Stan code and additional convergence and model checking results

Gaussian linear model with regularized horseshoe prior

```
p0 <- 5 # prior guess for the number of relevant variables
tau0 <- p0/(p-p0) * 1/sqrt(n)
rhs_prior <- hs(global_scale=tau0)
fitrhs <- stan_glm(formula, data = df, prior=rhs_prior,
                   QR=TRUE, seed=SEED, refresh=0)
```

For figures, e.g.,

```
theme_set(bayesplot::theme_default(base_family = "sans",
          base_size=16))
```