

# Modeling student retention

An example of project presentation slides

Aki Vehtari  
Aalto University

# Modeling student retention

An example of project presentation slides

Aki Vehtari  
Aalto University

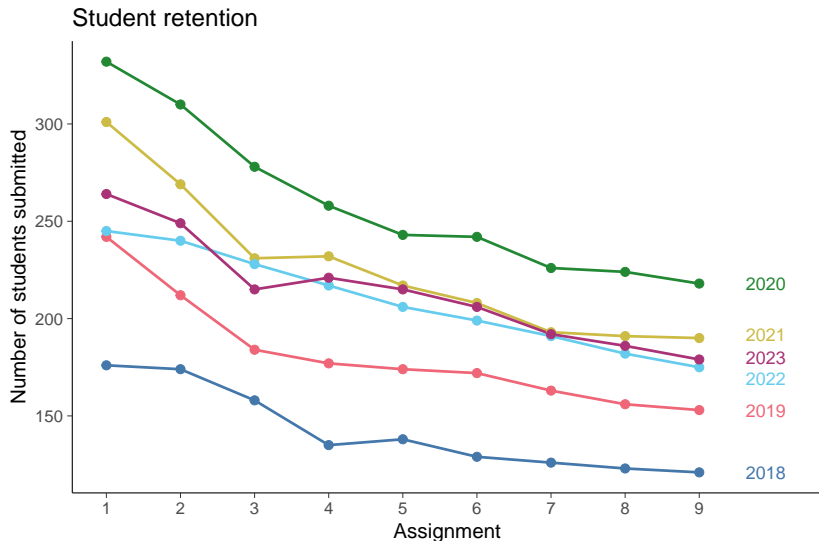
Introduce yourself

# Modeling student retention

- Is there difference in student retention in different years?
- When making changes to the course and assignment it is useful to follow changes in retention
  - although external effects (like pandemic) make it difficult to make precise conclusions

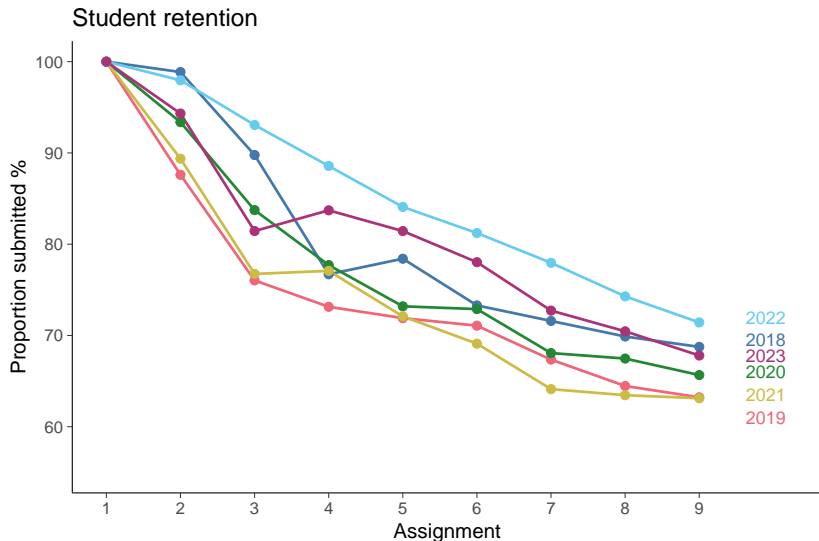
# Modeling student retention

Is there difference in student retention in different years?



# Modeling student retention

Is there difference in student retention in different years?



# Student retention

- As we want to compare different years, we use hierarchical models and compare

## 1. Latent hierarchical linear model

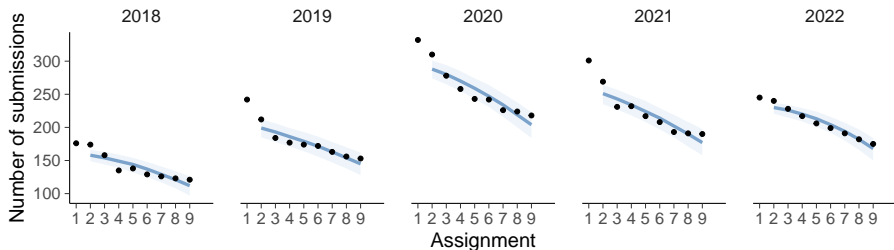
```
nstudents | trials(nstudents1) ~ (assignment | year),  
family=binomial()
```

## 2. Latent hierarchical linear model + spline

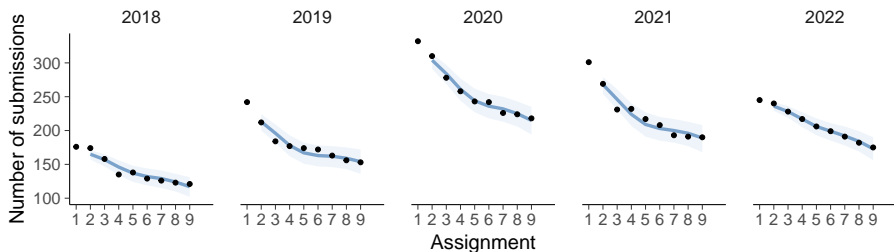
```
nstudents | trials(nstudents1) ~ (assignment | year) +  
s(assignment, k=4), family=binomial()
```

# Student retention – Posterior predictive distributions

## Latent hierarchical linear model

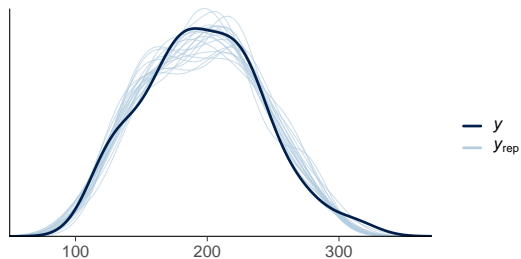


## Latent hierarchical linear model + spline

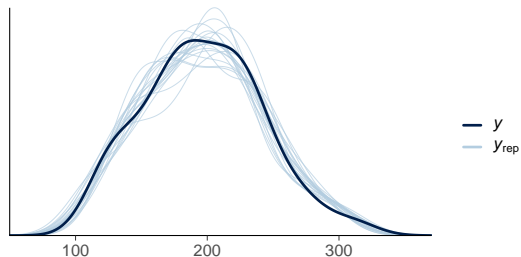


# Student retention – Marginal PPC

Latent hierarchical linear model



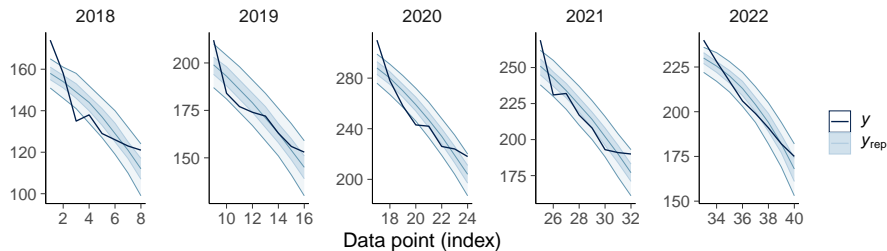
Latent hierarchical linear model + spline



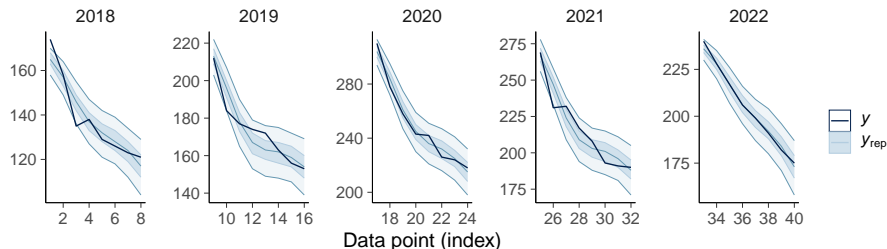


# Student retention – Posterior predictive ribbon

## Latent hierarchical linear model

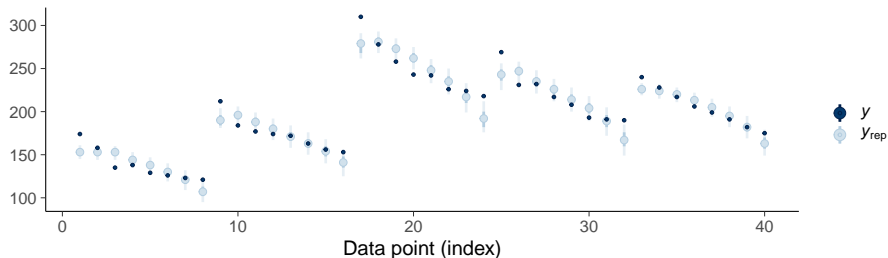


## Latent hierarchical linear model + spline

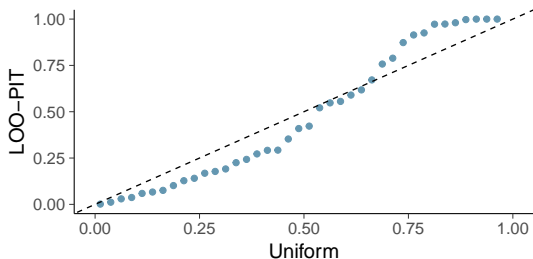


# Student retention – LOO-PIT checking

## Latent hierarchical linear – LOO predictive intervals

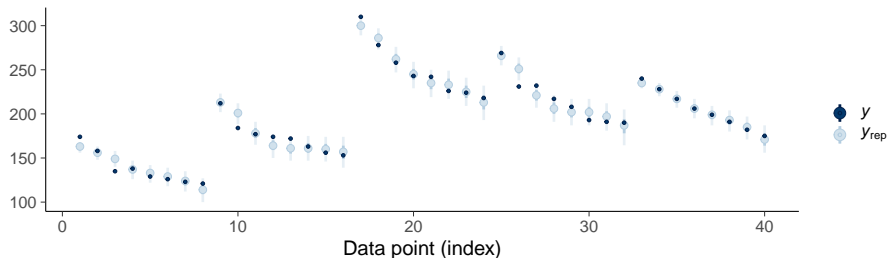


## LOO-PIT check

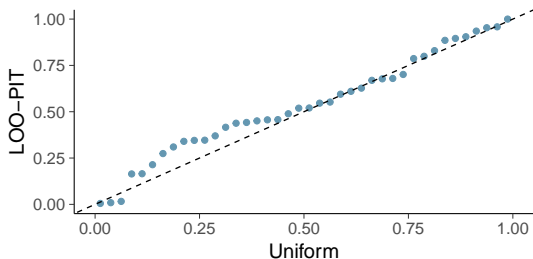


# Student retention – LOO-PIT checking

Latent hierarchical linear + spline – LOO predictive intervals/



LOO-PIT check

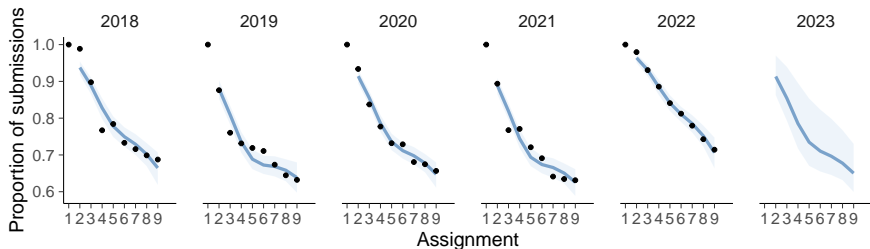


# Student retention – LOO model comparison

Latent hierarchical linear vs. latent hierarchical linear + spline

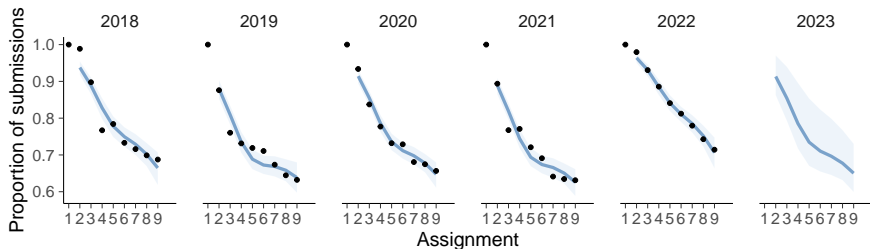
- $R^2$ : 0.92 vs 0.97 in favor of +spline
- ELPD-difference: 43 (SE 14) in favor of +spline

# Student retention latent spline model, year 2023?



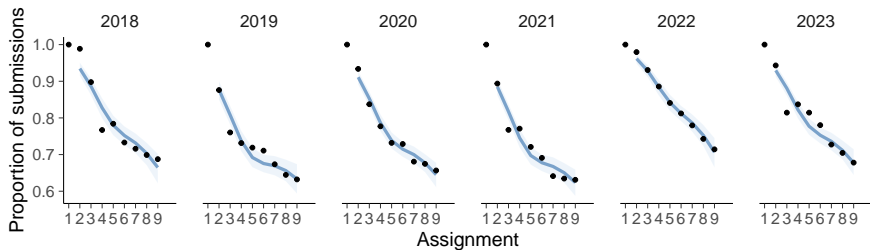
- 90% posterior predictive interval for assignment 9 (155, 193)

# Student retention latent spline model, year 2023?

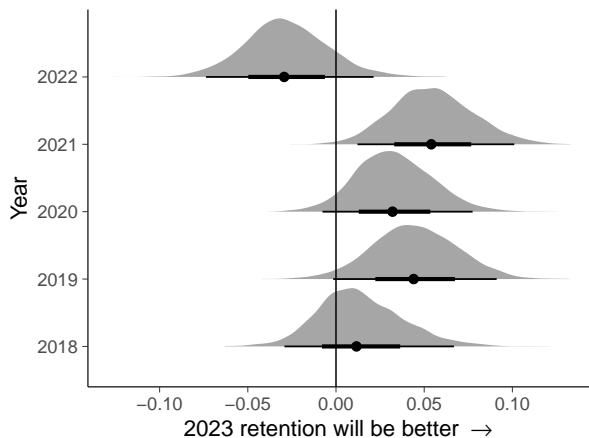


- 90% posterior predictive interval for assignment 9 (155, 193)
- Actually observed 179

# How does year 2023 compare to others?



# How does year 2023 compare to others?





# Convergence diagnostics

- MCMC sampling performed well based on the usual convergence diagnostics

# Conclusions

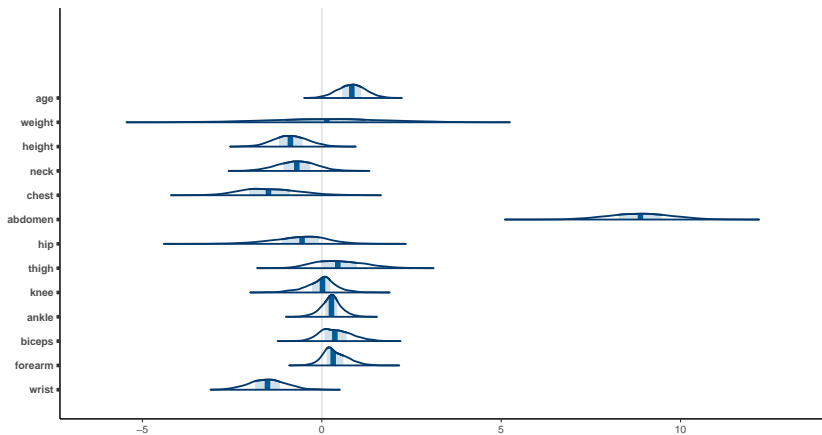
- Latent hierarchical model with spline can model 98% of the variation submission numbers
- Based on the model checking diagnostics, the model is reasonable
- Year 2023 retention was slightly above average

## Extra hints

- Following slides discuss some presentation hints

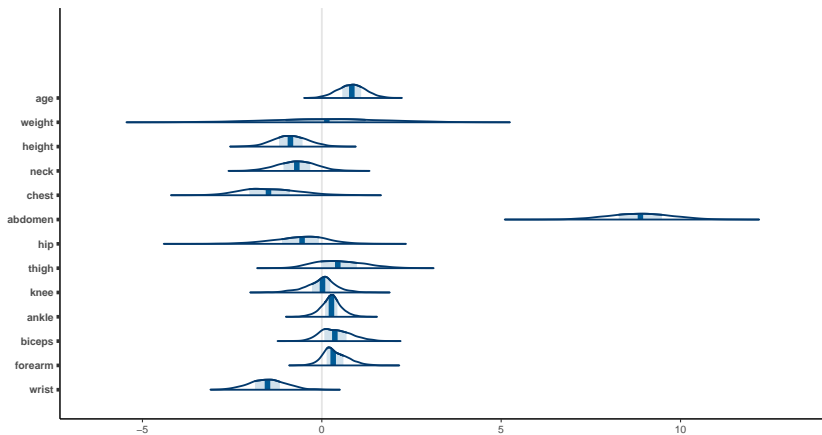
# Bodyfat

## Marginal posteriors of coefficients



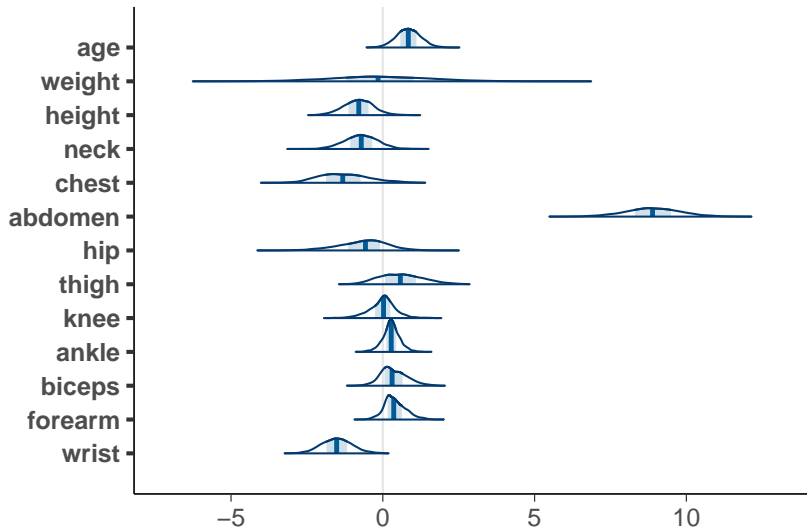
# Bodyfat

Check that the font in all figures is big enough!



# Bodyfat

Marginal posteriors of coefficients (Much better!)



## Figure font size

For example:

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

## Projective predictive covariate selection

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

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

where  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$  and  $\boldsymbol{\beta}$  is in general non-sparse (all  $\beta_j \neq 0$ )

- What is the best distribution  $q_{\perp}(\boldsymbol{\theta})$  given a constraint that only selected covariates have nonzero coefficient
- Optimization problem:

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

- 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  $\tilde{y}$

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

where  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$  and  $\boldsymbol{\beta}$  is in general non-sparse (all  $\beta_j \neq 0$ )

- What is the best distribution  $q_{\perp}(\boldsymbol{\theta})$  given a constraint that only selected covariates have nonzero coefficient
- Optimization problem:

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

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

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

# Conclusions

- Latent hierarchical model with spline can model 98% of the variation submission numbers
- Based on the model checking diagnostics, the model is reasonable
- Year 2023 retention was slightly above average

## 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

```
// generated with brms 2.14.4
functions {
  vector horseshoe(vector z, vector lambda, real tau, real c2) {
    int K = rows(z);
    vector[K] lambda2 = square(lambda);
    vector[K] lambda_tilde = sqrt(c2 * lambda2 ./ (c2 + tau^2 * lambda2));
    return z .* lambda_tilde * tau;
  }
}
data {
  int<lower=1> N; // total number of observations
  vector[N] Y; // response variable
  int<lower=1> K; // number of population-level effects
  matrix[N, K] X; // population-level design matrix
  // data for the horseshoe prior
  real<lower=0> hs_df; // local degrees of freedom
  real<lower=0> hs_df_global; // global degrees of freedom
  real<lower=0> hs_df_slab; // slab degrees of freedom
  real<lower=0> hs_scale_global; // global prior scale
  real<lower=0> hs_scale_slab; // slab prior scale
  int prior_only; // should the likelihood be ignored?
}
transformed_data {
```