

STATS 551 - HW 5 - Q3

Exercise - 3

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
library(ggplot2)
```

```
library(rstanarm)
```

Loading required package: Rcpp

This is rstanarm version 2.32.1

- See <https://mc-stan.org/rstanarm/articles/priors> for changes to default priors!
- Default priors may change, so it's safest to specify priors, even if equivalent to the defaults.
- For execution on a local, multicore CPU with excess RAM we recommend calling
options(mc.cores = parallel::detectCores())

```
library(loo)
```

This is loo version 2.8.0

- Online documentation and vignettes at mc-stan.org/loo
- As of v2.0.0 loo defaults to 1 core but we recommend using as many as possible. Use the 'cores' argument
- Windows 10 users: loo may be very slow if 'mc.cores' is set in your .Rprofile file (see <https://github.com/stan-dev/rstanarm>)

```
library(Rcpp)
```

```
options(mc.cores = parallel::detectCores())
```

```
d1 <- read.csv("student-mat.csv") # math results  
d2 <- read.csv("student-por.csv") # portugese results
```

```
# print(head(d1))
# print(head(d2))
#
# nrow(d1)
# nrow(d2)
#
# colnames(d1)
# colnames(d2)

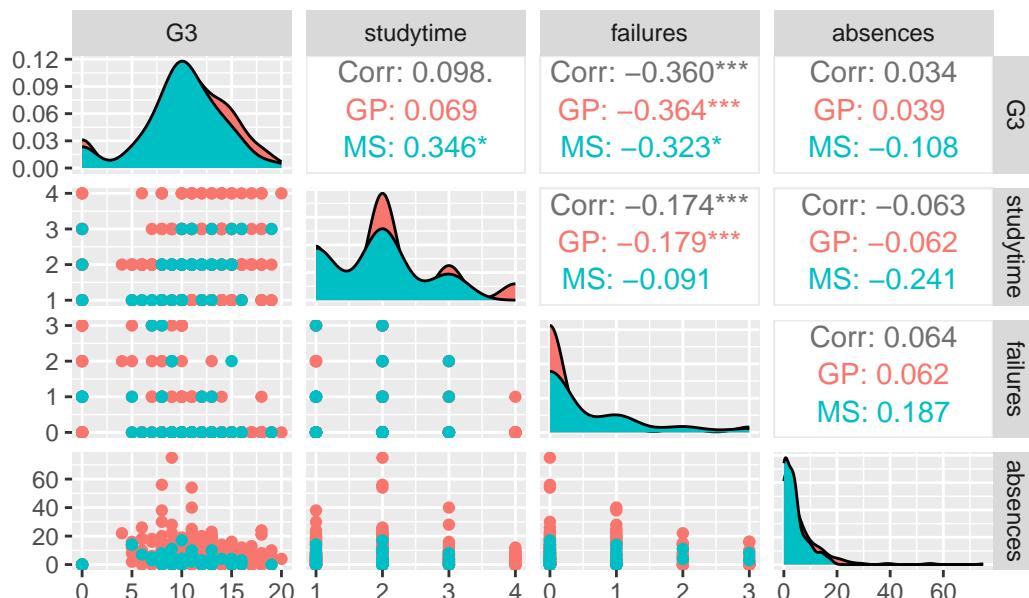
# Pairwise correlation
library(GGally)
```

Registered S3 method overwritten by 'GGally':

```
method from
+.gg   ggplot2
```

```
# Visualize relationships between continuous predictors and grade
ggpairs(d1, columns = c("G3", "studytime", "failures", "absences"),
         aes(color = school)) +
  labs(title = "Pairwise Correlation Between Predictors and Grade")
```

Pairwise Correlation Between Predictors and Grade



```
# Interaction between gender, studytime, and grade
ggplot(d1, aes(x = studytime, y = G3, color = sex)) +
  geom_smooth(method = "loess", se = FALSE) +
  labs(title = "Interaction Between Study Time and Grade by Gender",
       x = "Study Time", y = "Final Grade")
```

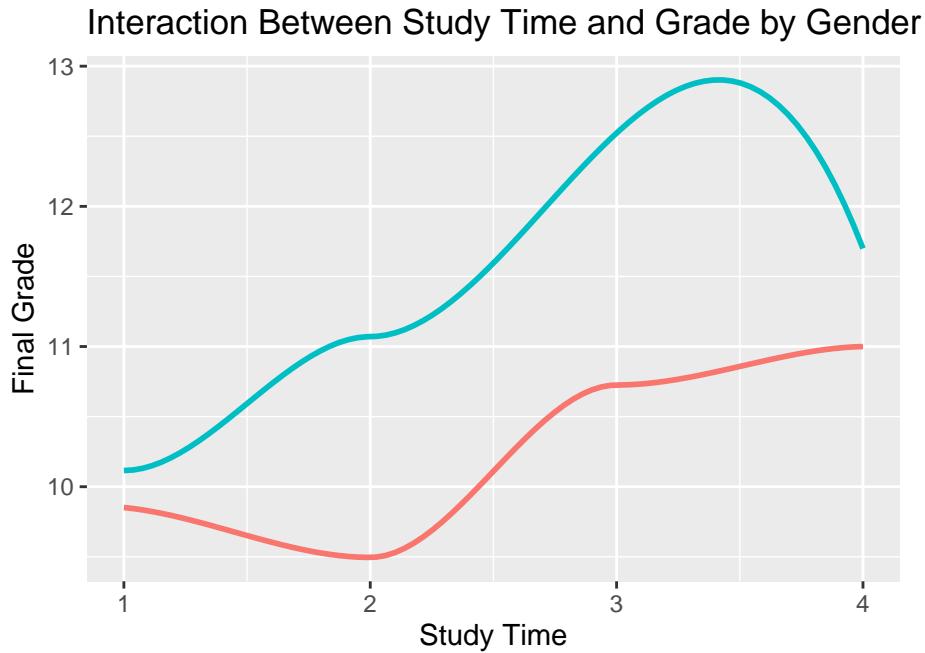
`geom_smooth()` using formula = 'y ~ x'

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric, : pseudoinverse used at 0.985

```

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 2.015
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 6.0033e-16
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 4.0602
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 0.985
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 1.015
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1.4607e-30
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 1

```



```

# Interaction model
interaction_model <- lm(G3 ~ studytime * sex + failures + absences, data = d1)

# Model summary
summary(interaction_model)

```

Call:
`lm(formula = G3 ~ studytime * sex + failures + absences, data = d1)`

Residuals:

Min	1Q	Median	3Q	Max
-12.5514	-1.9246	0.2956	2.9146	9.4412

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   9.28052   0.93682   9.906 < 2e-16 ***
studytime     0.47985   0.37384   1.284    0.200  
sexM          1.41074   1.17328   1.202    0.230  
failures      -2.19830  0.29330  -7.495 4.51e-13 ***
absences       0.04153   0.02690   1.544    0.123  
studytime:sexM -0.01480  0.53751  -0.028    0.978  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

Residual standard error: 4.238 on 389 degrees of freedom
Multiple R-squared:  0.1551,    Adjusted R-squared:  0.1442 
F-statistic: 14.28 on 5 and 389 DF,  p-value: 7.603e-13
```

```

# Hierarchical model with varying slopes
advanced_hierarchical_model <- stan_glmer(
  G3 ~ studytime + failures + absences + (0 + studytime | school) + (0 + 1 | sex),
  data = d1,
  family = gaussian(),
  prior_intercept = normal(10, 2),
  prior = normal(0, 1),
  prior_aux = cauchy(0, 2.5),
  chains = 4,
  iter = 4000,
  seed = 123,
  control = list(adapt_delta = 0.99,max_treedepth = 15)

)
```

Warning: There were 122 divergent transitions after warmup. See
<https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup>
 to find out why this is a problem and how to eliminate them.

Warning: Examine the pairs() plot to diagnose sampling problems

```
# Model summary
summary(advanced_hierarchical_model)
```

Model Info:

```

function:      stan_glmer
family:        gaussian [identity]
formula:       G3 ~ studytime + failures + absences + (0 + studytime | school) +
               (0 + 1 | sex)
algorithm:     sampling
sample:        8000 (posterior sample size)
priors:        see help('prior_summary')
observations:  395
```

```
groups: school (2), sex (2)
```

Estimates:

	mean	sd	10%	50%	90%
(Intercept)	9.9	1.3	8.3	9.9	11.4
studytime	0.4	0.5	-0.2	0.4	0.9
failures	-2.0	0.3	-2.4	-2.0	-1.7
absences	0.0	0.0	0.0	0.0	0.1
b[studytime school:GP]	0.1	0.4	-0.4	0.0	0.6
b[studytime school:MS]	0.2	0.5	-0.3	0.1	0.8
b[(Intercept) sex:F]	-0.5	1.2	-1.9	-0.5	0.8
b[(Intercept) sex:M]	0.7	1.2	-0.5	0.6	2.1
sigma	4.2	0.2	4.0	4.2	4.4
Sigma[school:studytime,studytime]	2.4	9.8	0.0	0.2	4.0
Sigma[sex:(Intercept),(Intercept)]	8.5	24.0	0.3	2.2	18.6

Fit Diagnostics:

	mean	sd	10%	50%	90%
mean_PPD	10.4	0.3	10.0	10.4	10.8

The mean_ppd is the sample average posterior predictive distribution of the outcome variable (for detailed information see the Stan manual).

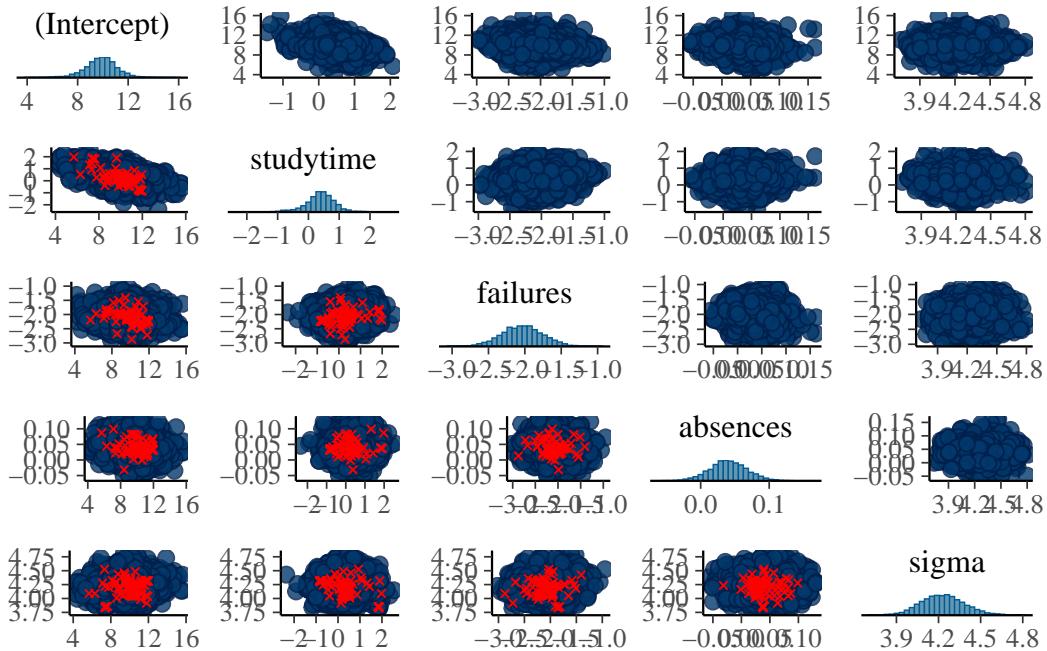
MCMC diagnostics

	mcse	Rhat	n_eff
(Intercept)	0.0	1.0	3506
studytime	0.0	1.0	1854
failures	0.0	1.0	7354
absences	0.0	1.0	7018
b[studytime school:GP]	0.0	1.0	1651
b[studytime school:MS]	0.0	1.0	1343
b[(Intercept) sex:F]	0.0	1.0	3018
b[(Intercept) sex:M]	0.0	1.0	2945
sigma	0.0	1.0	5406
Sigma[school:studytime,studytime]	0.7	1.0	190
Sigma[sex:(Intercept),(Intercept)]	0.7	1.0	1046
mean_PPD	0.0	1.0	7697
log-posterior	0.1	1.0	1853

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample size.

```
pairs(advanced_hierarchical_model,
      pars = c("(Intercept)", "studytime", "failures", "absences", "sigma"),
      pch = 20,
      cex = 0.4)
```

Warning: The following arguments were unrecognized and ignored: pch, cex



```
summary(advanced_hierarchical_model)
```

Model Info:

```
function:      stan_glmer
family:        gaussian [identity]
formula:       G3 ~ studytime + failures + absences + (0 + studytime | school) +
               (0 + 1 | sex)
algorithm:    sampling
sample:        8000 (posterior sample size)
priors:        see help('prior_summary')
observations: 395
groups:        school (2), sex (2)
```

Estimates:

	mean	sd	10%	50%	90%
(Intercept)	9.9	1.3	8.3	9.9	11.4
studytime	0.4	0.5	-0.2	0.4	0.9
failures	-2.0	0.3	-2.4	-2.0	-1.7
absences	0.0	0.0	0.0	0.0	0.1
b[studytime school:GP]	0.1	0.4	-0.4	0.0	0.6
b[studytime school:MS]	0.2	0.5	-0.3	0.1	0.8
b[(Intercept) sex:F]	-0.5	1.2	-1.9	-0.5	0.8
b[(Intercept) sex:M]	0.7	1.2	-0.5	0.6	2.1
sigma	4.2	0.2	4.0	4.2	4.4
Sigma[school:studytime,studytime]	2.4	9.8	0.0	0.2	4.0
Sigma[sex:(Intercept),(Intercept)]	8.5	24.0	0.3	2.2	18.6

Fit Diagnostics:

```

mean     sd    10%   50%   90%
mean_PPD 10.4     0.3 10.0  10.4  10.8

```

The mean_ppd is the sample average posterior predictive distribution of the outcome variable (for detailed explanation see [here](#))

MCMC diagnostics

	mcse	Rhat	n_eff
(Intercept)	0.0	1.0	3506
studytime	0.0	1.0	1854
failures	0.0	1.0	7354
absences	0.0	1.0	7018
b[studytime school:GP]	0.0	1.0	1651
b[studytime school:MS]	0.0	1.0	1343
b[(Intercept) sex:F]	0.0	1.0	3018
b[(Intercept) sex:M]	0.0	1.0	2945
sigma	0.0	1.0	5406
Sigma[school:studytime,studytime]	0.7	1.0	190
Sigma[sex:(Intercept),(Intercept)]	0.7	1.0	1046
mean_PPD	0.0	1.0	7697
log-posterior	0.1	1.0	1853

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample size.

```

posterior_summary <- as.data.frame(posterior_interval(advanced_hierarchical_model))
print(posterior_summary)

```

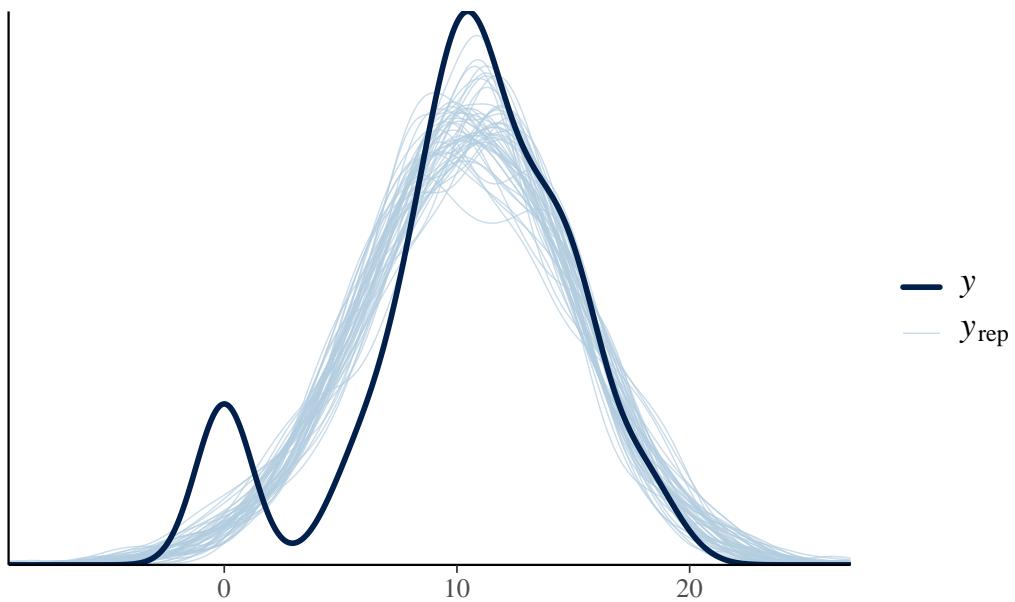
	5%	95%
(Intercept)	7.684803190	11.86183401
studytime	-0.456541709	1.08684677
failures	-2.498991945	-1.56764640
absences	-0.004339904	0.08580278
b[studytime school:GP]	-0.557159194	0.86625736
b[studytime school:MS]	-0.463538850	1.08573242
b[(Intercept) sex:F]	-2.401351131	1.36718496
b[(Intercept) sex:M]	-1.063535685	2.70176240
sigma	3.989168571	4.49323821
Sigma[school:studytime,studytime]	0.001263832	8.78707645
Sigma[sex:(Intercept),(Intercept)]	0.170298758	33.10170253

```

# Global PPC
pp_check(advanced_hierarchical_model) +
  ggtitle("Global Posterior Predictive Check for Hierarchical Model")

```

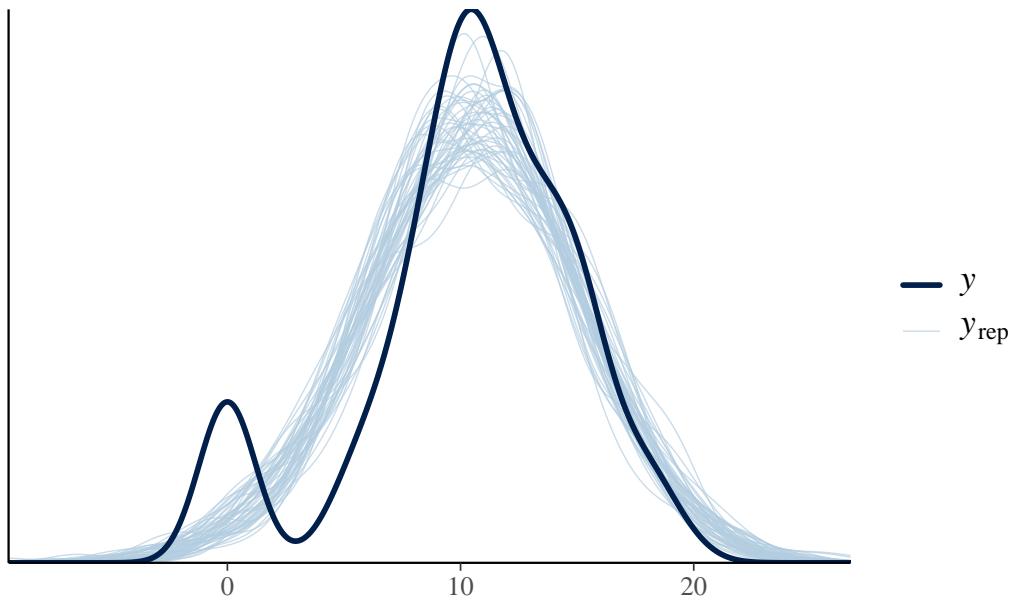
Global Posterior Predictive Check for Hierarchical Model



```
# Group-level PPC
pp_check(advanced_hierarchical_model, group = "school") +
  ggtitle("Posterior Predictive Check by School")
```

Warning: The following arguments were unrecognized and ignored: group

Posterior Predictive Check by School



```
# library(shinystan)
# launch_shinystan(advanced_hierarchical_model)

# posterior <- as.matrix(advanced_hierarchical_model)
# divergences <- sum(attr(posterior, "sampler_params")$divergent__)
# print(divergences)
```

```

# Install and load rstanarm if needed
if (!requireNamespace("rstanarm", quietly = TRUE)) install.packages("rstanarm")
library(rstanarm)

# Bayesian interaction model
bayesian_interaction_model <- stan_glm(
  G3 ~ studytime * sex + failures + absences,
  data = d1,
  family = gaussian(),
  prior_intercept = normal(10, 5),
  prior = normal(0, 2.5),
  prior_aux = cauchy(0, 2),
  chains = 4,
  iter = 2000,
  seed = 123
)

```

```

# Compute WAIC for the Bayesian models
waic_one_param <- waic(bayesian_interaction_model)

```

Warning:

1 (0.3%) p_waic estimates greater than 0.4. We recommend trying loo instead.

```

waic_hierarchical <- waic(advanced_hierarchical_model)

# Print WAIC results
print(waic_one_param)

```

Computed from 4000 by 395 log-likelihood matrix.

	Estimate	SE
elpd_waic	-1135.0	15.7
p_waic	7.2	0.8
waic	2270.0	31.4

1 (0.3%) p_waic estimates greater than 0.4. We recommend trying loo instead.

```
print(waic_hierarchical)
```

Computed from 8000 by 395 log-likelihood matrix.

	Estimate	SE
elpd_waic	-1134.9	15.7
p_waic	7.1	0.7
waic	2269.9	31.4

```
# Compute LOO for both models
loo_one_param <- loo(bayesian_interaction_model)
```

```

loo_hierarchical <- loo(advanced_hierarchical_model)

# Compare models using loo_compare
loo_comparison <- loo_compare(loo_one_param, loo_hierarchical)

# Print LOO comparison results
print(loo_comparison)

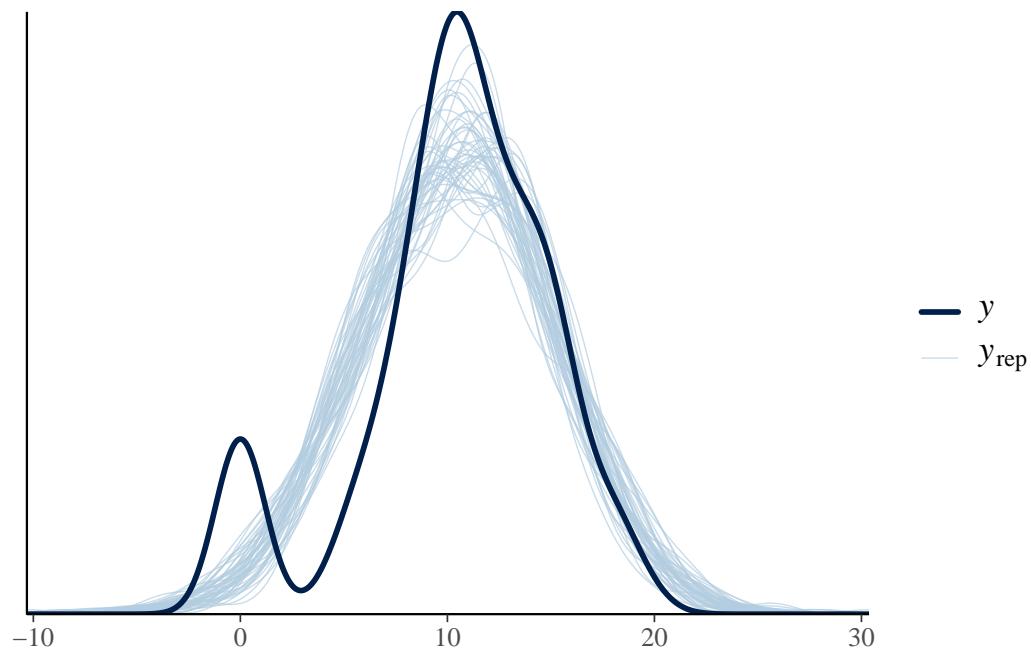
```

	elpd_diff	se_diff
advanced_hierarchical_model	0.0	0.0
bayesian_interaction_model	-0.1	0.8

```

# Posterior predictive checks for the interaction model
pp_check(bayesian_interaction_model)

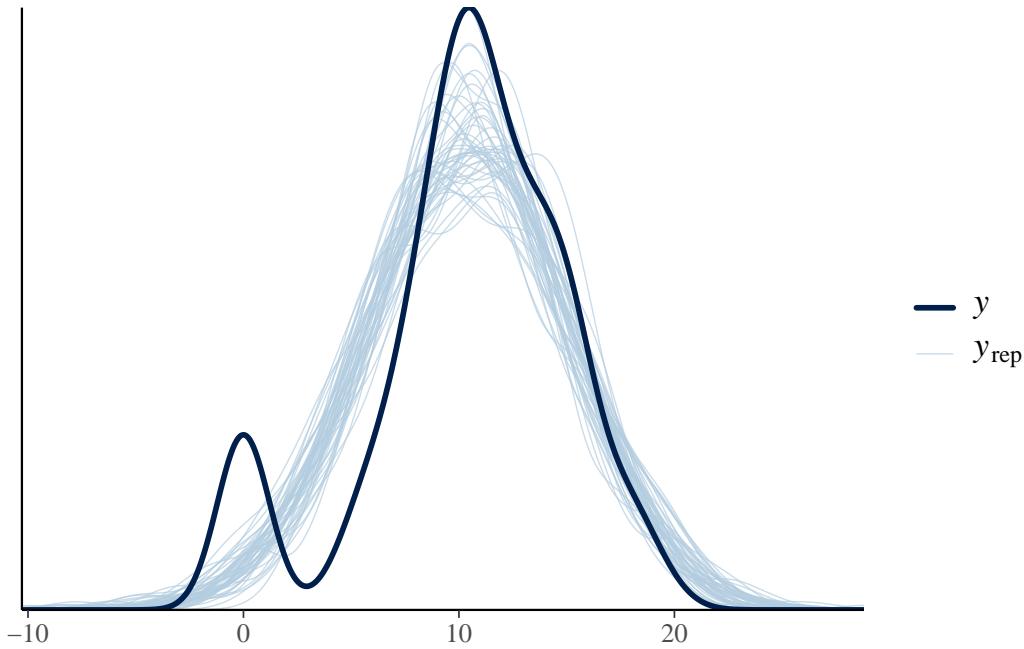
```



```

# Posterior predictive checks for the hierarchical model
pp_check(advanced_hierarchical_model)

```



```
# Extract posterior samples for random effects
library(bayesplot)
```

This is bayesplot version 1.11.1

- Online documentation and vignettes at mc-stan.org/bayesplot
- bayesplot theme set to `bayesplot::theme_default()`
 - * Does not affect other `ggplot2` plots
 - * See `?bayesplot_theme_set` for details on theme setting

```
posterior_samples <- as.data.frame(advanced_hierarchical_model)
```

```
colnames(posterior_samples) # List all parameter names
```

```
[1] "(Intercept)"                      "studytime"
[3] "failures"                         "absences"
[5] "b[studytime school:GP]"           "b[studytime school:MS]"
[7] "b[(Intercept) sex:F]"            "b[(Intercept) sex:M]"
[9] "sigma"                            "Sigma[school:studytime,studytime]"
[11] "Sigma[sex:(Intercept),(Intercept)]"
```

```
# Visualize posterior for group-level intercepts (schools)
```

```
mcmc_areas(
  posterior_samples,
  pars = c("b[studytime school:GP]", "b[studytime school:MS]"),
  prob = 0.8
) +
  labs(
    title = "Posterior Distributions for School-Level Effects",
    x = "Effect Size",
```

```
y = "Density"  
)
```

Posterior Distributions for School–Level Effects

