

BDA - Assignment 7

Anonymous

Contents

Load packages	2
Exercise 1	2
Exercise 2	6

Load packages

```
library(aaltobda)
library("rstan")
library(magrittr)
library(tidyr)

library(loo)
library(ggplot2)
library(gridExtra)
library(bayesplot)
theme_set(bayesplot::theme_default(base_family = "sans"))
#library(shinystan)
library(rprojroot)

SEED <- 48927 # set random seed for reproducibility

data("drowning")
```

Exercise 1

1

The three errors are represented in the following image:

```
1 data {
2   int<lower=0> N; // number of data points
3   vector[N] x; // observation year
4   vector[N] y; // observation number of drowned
5   real xpred; // prediction year
6 }
7 parameters {
8   real alpha;
9   real beta;
10  real<upper=0> sigma;
11 }
12 transformed parameters {
13   vector[N] mu = alpha + beta*x;
14 }
15 model {
16   y ~ normal(mu, sigma);
17 }
18 generated quantities {
19   real ypred = normal_rng(mu, sigma);
20 }
```

Figure 1: Representation of the mistakes.

- There is a mistake in the definition of the standard deviation, sigma, in line 10. It must be constrained to be positive. However, in the original code, it is constrained to be smaller than 0. Hence, in order to

solve the error, it must be defined as: `real < lower=0 > sigma;`

- A syntax mistake can be found in line 16. Every stan code line must end with a semicolon. Hence, to solve the error, it is enough to add it: `y ~ normal(mu, sigma);`
- Finally, there is another mistake in the definition of the predicted samples, line 19. When establishing the parameters of the normal distribution for the simulated samples, the mean must depend on the prediction year. Instead, it was defined depending on the observation one. In order to solve it: `real ypred = normal_rng(alpha+beta*xpred, sigma);`

The fixed code is:

```
data {
  int < lower =0 > N ; // number of data points
  vector [ N ] x ;
  // observation year
  vector [ N ] y ;
  // observation number of drowned
  real xpred ;
  // prediction year
  real pmualpha; // prior mean for alpha
  real psalpha; // prior std for alpha
  real pmubeta; // prior mean for beta
  real psbeta; // prior std for beta
}
parameters {
  real alpha ;
  real beta ;
  real < lower =0 > sigma ;
}
transformed parameters {
  vector [ N ] mu = alpha + beta * x ;
}
model {
  alpha ~ normal(pmualpha, psalpha); // prior
  beta ~ normal(pmubeta, psbeta); // prior
  y ~ normal ( mu , sigma );
}
generated quantities {
  real ypred = normal_rng ( alpha+beta*xpred, sigma );
}
```

2

In order to compute the standard deviation of the weakly-informative prior for the slope beta, the calculation can be based on the quantiles. Knowing that $Pr(-69 < beta < 69) = 0.99$, then, the function `qnorm` can be used to find out what value of the standard deviation yields -69 and 69. Since:

$$Pr(-69 < beta < 69) = Pr(beta < 69) - Pr(beta < -69) = 0.99$$

Then, 69 corresponds with the quantil 99.5%, while -69 represents the quantil 0.5%. Different standard deviations are tried until the values 69 and -69 are obtained.

```
qnorm(c(0.005,0.995), 0, 27)
```

```
## [1] -69.54739 69.54739
```

The obtained value, 27, is a reasonable value. According to the plot of a normal distribution, as it is shown in the figure 2, it makes sense to obtain a value higher than 23, since:

$$\sigma_\beta > \frac{69 - -69}{6} = 23$$

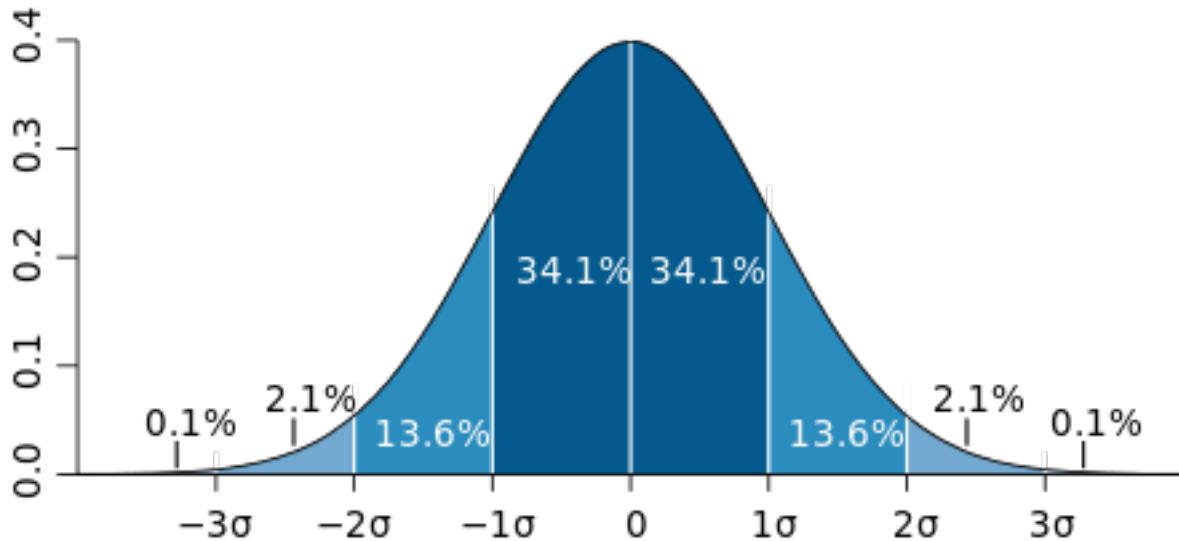


Figure 2: Representation of the normal distribution and the standard deviation.

Hence, the chosen value is:

$$\sigma_\beta = 27$$

3

In order to implement in the code the prior, inside the block corresponding to model, the following line was added:

- `beta ~ normal(pmubeta, psbeta);`

Where, pmubeta is 0, as specified in the exercise, and psbeta is the previously obtained value $\sigma_\beta = 27$.

4

In order to add a weakly informative prior for the intercept alpha, another normal distribution with zero mean can be used. The standard deviation must be a high value, in order to make it weakly informative, somehow related to the context of the problem. As stated in the book BDA3, *in the general problem of estimating a normal mean, a $N(0, A^2)$ prior distribution is weakly informative, with A set to some large value that depends on the context of the problem*. The chosen value for σ_α is the square mean of the drownings:

```
mean(drowning$drownings)
```

```
## [1] 134.35
```

Hence, the chosen value is:

$$\sigma_\alpha = 135^2$$

In the figure 3, the plots obtained from the execution are shown:

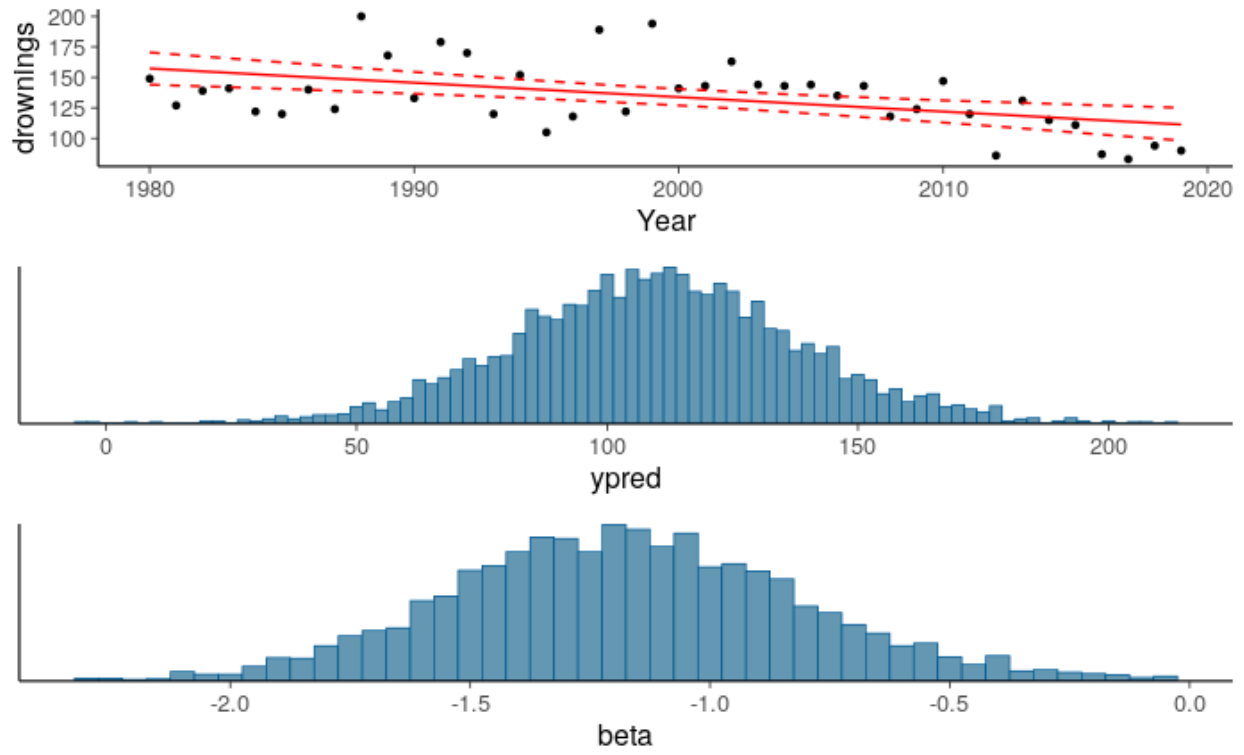


Figure 3: Obtained graphs (the predictive histogram is for the year 2020).

Exercise 2

```
data("factory")
```

Separate model

In the separate model, each machine has its own model, meaning that they have different standard deviation σ_j for each group j . The priors used are those suggested in the last part of this exercise, since they are weakly informative.

a

In this initial part of the exercise, the entire code is included. However, it is not for the next questions (hierarchical and pooled models), since it is analogous.

The data required for running the model is defined and the model is implemented, yielding the following results for the plots:

The code for obtaining the required plots is:

```
draws_separate <- rstan::extract(separate_sampling, permuted = T)

pars_separate <- intersect(names(draws_separate), c('ypred'))
draws_separate <- as.data.frame(separate_sampling)
phist_separate <- mcmc_hist(draws_separate, pars = c(pars_separate,
                                                    names(draws_separate)[6]),
                           binwidth = 2.5)

grid.arrange(phist_separate, nrow = 1)
```

In the figure 4, the plots for answering the questions are included.

i The posterior distribution of the mean represents the prior used for computing the likelihood. In this case, the histogram seems reasonable considering the values of the data set. It has the shape of a normal distribution with a mean around the value 50 and a standard deviation that could be around 10.

ii The predictive distribution for another quality measure of the sixth machine represents the distribution that follows a simulated sample for the sixth machine. It is implemented in the model block “generated quantities”. In the given case, it is highly probable that the next measure would be around 50, since the distribution is normal with mean around the mentioned value.

iii Since comparison between k groups is being performed, there is no possibility to extend the knowledge to a seventh machine. Instead, it is possible to obtain simulations of the future quality measures of one of the machines, as it was done in the previous question.

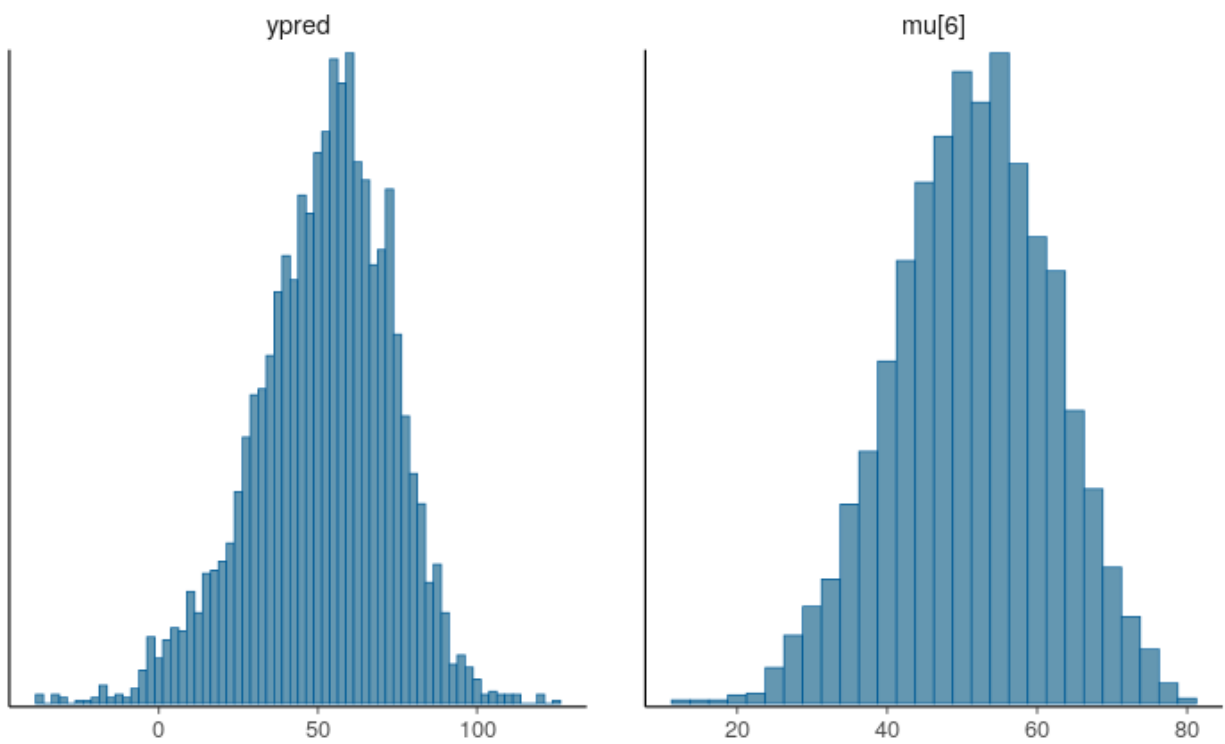


Figure 4: Left: predictive distribution for another quality measurement of the sixth machine. Right: posterior distribution of the mean of the quality measurements of the sixth machine.

b

The code for the Stan model is:

```
data {
  int < lower =0 > N ;
  int < lower =0 > J ;
  vector [ J ] y [ N ];
  int p_mu;
  int < lower =0 > p_alpha;
  int < lower =0 > p_beta;
}
parameters {
  vector [ J ] mu ;
  vector < lower =0 >[ J ] sigma ;
}
model {
  // priors
  for ( j in 1: J ){
    mu [ j ] ~ normal ( 0 , p_mu);
    sigma [ j ] ~ gamma (p_alpha, p_beta);
  }
  // likelihood
  for ( j in 1: J )
    y [ , j ] ~ normal ( mu [ j ] , sigma [ j ]);
}
generated quantities {
  real ypred ;
  // Compute predictive distribution for the sixth machine
  ypred = normal_rng ( mu [ 6 ] , sigma [ 6 ]);
}
```

As priors, the suggested ones were used, a Normal distribution for the mean and a Gamma distribution for the standard deviation. Since it is a separate model, both priors are included in the for loop, yielding different results for each of the groups that are considered (machines).

c

The mathematical formulation of the model is:

$$y_{ij} \sim N(\mu_j, \sigma_j)$$

$$\mu_j \sim N(0, 10)$$

$$\sigma_j \sim \text{Gamma}(1, 1)$$

d

The posterior expectation for μ_1 can be obtained as:

```
monitor(separate_sampling)
```

```
## Inference for the input samples (4 chains: each with iter = 2000; warmup = 0):
##
##           Q5      Q50      Q95      Mean      SD      Rhat Bulk_ESS Tail_ESS
```



```

## mu[1]      35.3   51.3   64.1   50.6   8.8       1    3150    3076
## mu[2]      29.7   48.6   69.6   49.0  12.3       1    3557    2352
## mu[3]      38.9   59.1   76.4   58.7  11.4       1    3356    3291
## mu[4]      28.0   47.2   67.8   47.5  12.1       1    2756    1873
## mu[5]      38.7   61.8   80.5   60.9  12.9       1    3164    2941
## mu[6]      32.7   51.5   67.9   51.1  10.7       1    3487    2966
## sigma[1]   11.0   15.6   22.1   15.9   3.4       1    2837    3301
## sigma[2]   17.3   24.5   32.3   24.6   4.5       1    3063    2275
## sigma[3]    9.3   15.7   23.5   16.0   4.4       1    3250    3205
## sigma[4]   19.3   26.4   34.0   26.4   4.4       1    2694    1836
## sigma[5]    8.4   15.2   24.2   15.6   4.9       1    3196    3064
## sigma[6]   12.8   18.5   25.5   18.7   3.9       1    3253    2739
## ypred      12.3   52.8   85.1   51.5  22.1       1    3656    3774
## lp__       -351.3 -346.6 -343.6 -347.0   2.4       1    1757    2095
##
## For each parameter, Bulk_ESS and Tail_ESS are crude measures of
## effective sample size for bulk and tail quantities respectively (an ESS > 100
## per chain is considered good), and Rhat is the potential scale reduction
## factor on rank normalized split chains (at convergence, Rhat <= 1.05).

```

The 90% interval is defined by the Q95 and the Q5, hence, in the given case for the separate model and μ_1 would be: [34.8, 64.3].

Hierarchical

In a hierarchical model, a multilevel modeling process is considered, assuming that the prior distributions depend on an upon further hyperparameters that are obtained from a hyperprior distributions.

a

In the figure 5 there is the representation of the plots used for answering the questions:

i The posterior distribution of the mean of the quality measurements for the sixth machine follows a normal distribution with mean around 85 and standard deviation around 8. The hierarchical model is used so the population is not described by only one set of parameters, since another level is added. Hence, the obtained distribution should be more realistic than in the previous case.

ii The predictive distribution for another quality measurement follows a normal distribution with mean around 90 and standard deviation around 20. It is similar to the one obtained in the previous case but slightly displaced towards the right.

iii Since comparison between k groups is being performed, there is no possibility to extend the knowledge to a seventh machine. Instead, it is possible to obtain simulations of the future quality measures of one of the machines, as it was done in the previous question.

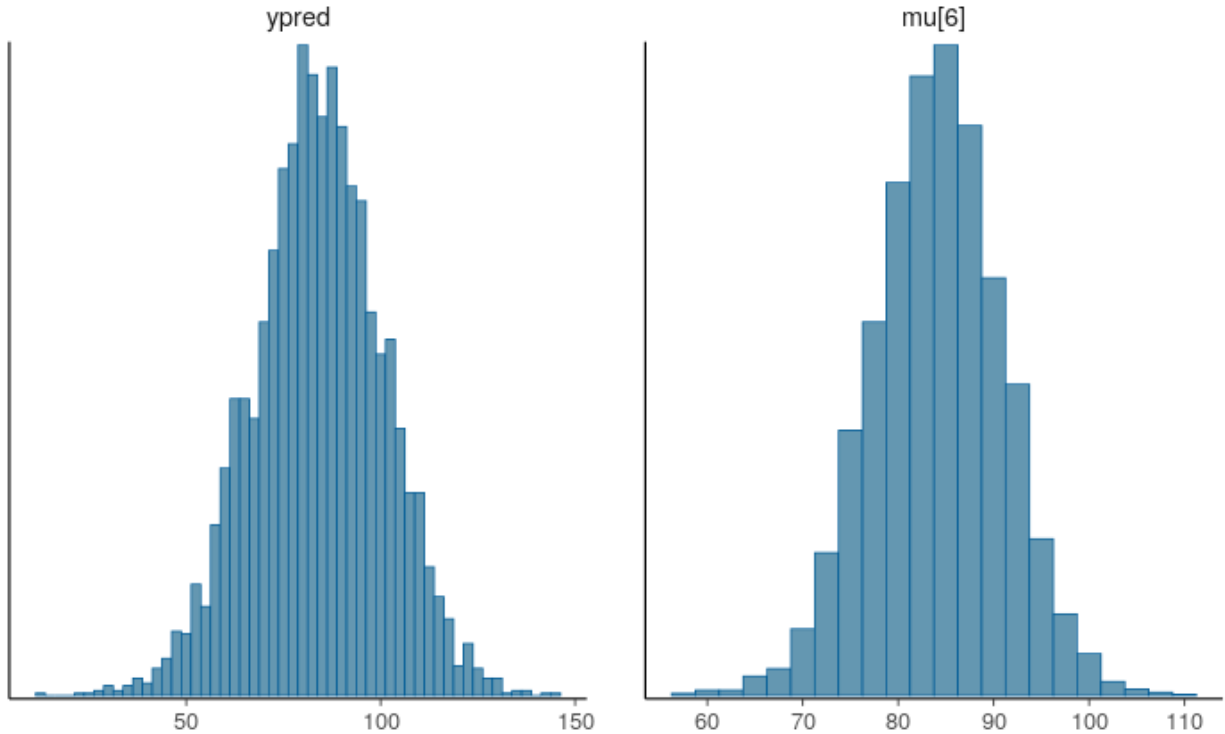


Figure 5: Left: predictive distribution for another quality measurement of the sixth machine. Right: posterior distribution of the mean of the quality measurements of the sixth machine.

b

The code for the model is:

```
data {
  int < lower =0 > N ;
  int < lower =0 > J ;
  vector [ J ] y [ N ];
  int < lower =0 > p_mu;
  int < lower =0 > p_alpha;
  int < lower =0 > p_beta;
}
parameters {
  vector [ J ] mu ;
  real tau;
  real theta;
  real sigma ;
}
model {
  // Hyperpriors
  tau ~ normal(0, p_mu);
  theta ~ gamma(p_alpha, p_beta);

  // Priors
  for (j in 1: J ){
    mu[ j ] ~ normal (0, tau);
  }
  sigma ~ inv_chi_square (theta);
  // likelihood
  for ( j in 1: J )
    y [ , j ] ~ normal ( mu [ j ] , sigma);
}
generated quantities {
  real ypred ;
  // Compute predictive distribution for the sixth machine
  ypred = normal_rng ( mu [6] , sigma);
}
```

c

The model expressed in mathematical notation is:

$$\begin{aligned}y_{ij} &\sim N(\mu_j, \sigma) \\ \textit{Hyperpriors} : \\ \tau &\sim N(0, 10) \\ \theta &\sim \textit{Gamma}(1, 1) \\ \textit{Priors} : \\ \mu_j &\sim N(0, \tau) \\ \sigma &\sim \textit{inv} - X^2(\theta)\end{aligned}$$

d

The posterior expectation for μ_1 can be obtained as:

```
monitor(hierarchical_sampling)
```

```
## Inference for the input samples (4 chains: each with iter = 2000; warmup = 0):
```

```
##
```

	Q5	Q50	Q95	Mean	SD	Rhat	Bulk_ESS	Tail_ESS
## mu[1]	63.2	74.4	85.5	74.5	6.8	1.00	2992	2295
## mu[2]	92.2	103.8	114.3	103.6	6.8	1.00	2947	2025
## mu[3]	74.8	85.9	96.7	85.8	6.9	1.00	2874	1887
## mu[4]	97.8	109.1	119.8	109.0	6.8	1.00	2645	1885
## mu[5]	76.8	87.7	98.8	87.8	6.7	1.00	2829	2167
## mu[6]	72.9	84.0	94.6	84.0	6.7	1.00	2793	2030
## tau	37.8	44.8	53.7	45.1	4.9	1.00	2544	2143
## theta	0.1	0.6	1.7	0.7	0.5	1.01	1447	1412
## sigma	11.6	14.7	19.1	14.9	2.4	1.00	1634	1195
## ypred	56.9	83.8	110.4	83.9	16.4	1.00	3985	3553
## lp__	-152.1	-147.3	-144.3	-147.7	2.5	1.00	1075	1092

```
##
```

```
## For each parameter, Bulk_ESS and Tail_ESS are crude measures of
## effective sample size for bulk and tail quantities respectively (an ESS > 100
## per chain is considered good), and Rhat is the potential scale reduction
## factor on rank normalized split chains (at convergence, Rhat <= 1.05).
```

The posterior expectation for μ_1 with a 90% credibility interval is given by [63.2, 84.9].

Pooled

In the pooled model, all measurements are combined and no distinction is made between machines.

a

In the figure 6, the necessary plots for answering the questions are included.

i In the given case, it does not make sense to differentiate between the existing machines, since all the data is treated as if it was coming from only one machine. Hence, the posterior distribution of the only considered machine is shown in the figure 6. It follows a normal distribution with mean around 85, similar to the previous case.

ii The predictive distribution for another quality measure is also computed for the “general” machine that the pooled model considers, instead of the sixth one. It yields a normal distribution with mean around 80.

iii Since comparison between k groups is being performed, there is no possibility to extend the knowledge to a seventh machine. Instead, it is possible to obtain simulations of the future quality measures of one of the machines, as it was done in the previous question.

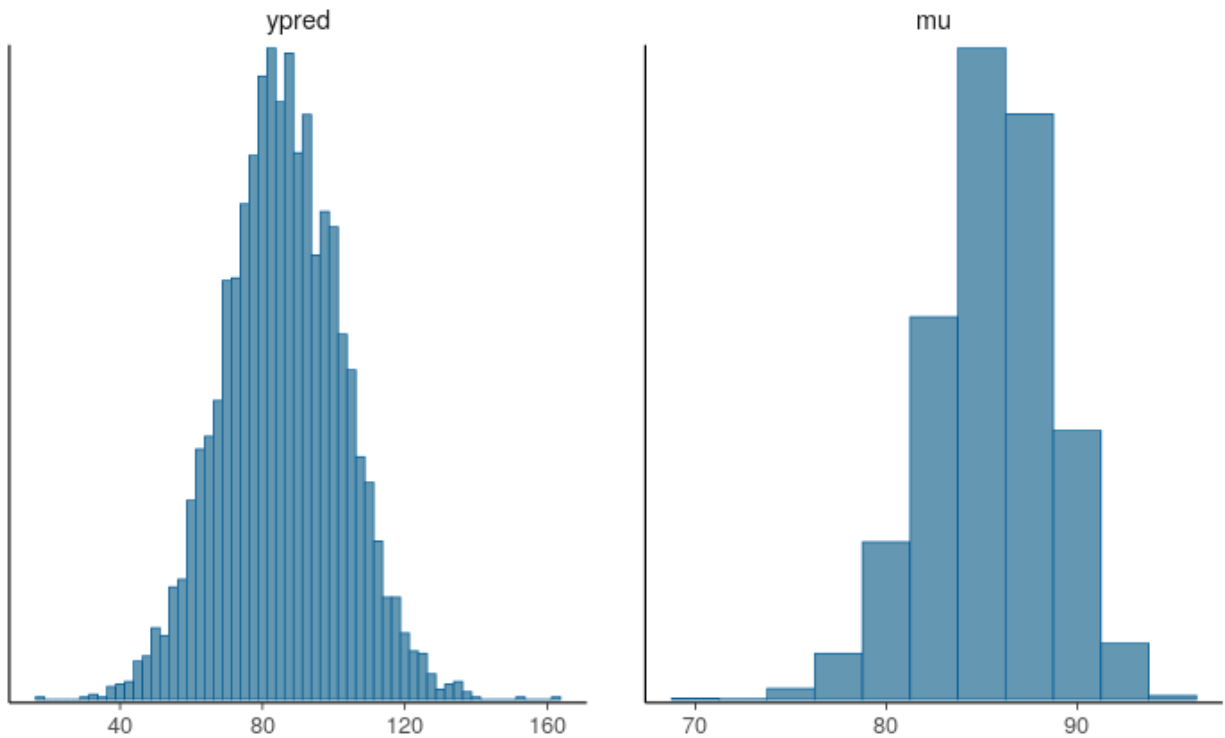


Figure 6: Left: predictive distribution for another quality measurement of the “general machine”. Right: posterior distribution of the mean of the quality measurements “general machine”.

b

The code for the stan model is:

```
data {
  int < lower =0 > N ;
  int < lower =0 > J ;
  vector [ J ] y [ N ];
  int < lower =0 > p_mu;
  int < lower =0 > p_alpha;
  int < lower =0 > p_beta;
}
parameters {
  real mu ;
  real sigma ;
}
model {
  // priors
  mu ~ normal (0 , p_mu);
  sigma ~ gamma (p_alpha, p_beta);
  // likelihood
  for ( j in 1: J )
    y [ , j ] ~ normal ( mu , sigma );
}
generated quantities {
  real ypred ;
  // Compute predictive distribution for the first machine
  ypred = normal_rng ( mu, sigma);
}
```

c

The model, described in mathematical notation is:

$$y \sim N(\mu, \sigma)$$

$$\mu \sim N(0, 10)$$

$$\sigma \sim \text{Gamma}(1, 1)$$

d

Finally, the posterior expectation for μ is computed from the 90% credibility interval:

```
monitor(pooled_sampling)
```

```
## Inference for the input samples (4 chains: each with iter = 2000; warmup = 0):
##
##           Q5    Q50    Q95   Mean   SD  Rhat Bulk_ESS Tail_ESS
## mu       80.0   85.7   90.5   85.5   3.2  1.00    1589    1723
## sigma    13.3   15.8   19.4   16.0   1.9  1.01    1581    1668
## ypred    57.8   86.2  112.8   85.7  16.6  1.00    3849    3577
## lp__   -160.5 -158.0 -157.3 -158.3  1.1  1.00    1479    1783
##
## For each parameter, Bulk_ESS and Tail_ESS are crude measures of
```

```
## effective sample size for bulk and tail quantities respectively (an ESS > 100
## per chain is considered good), and Rhat is the potential scale reduction
## factor on rank normalized split chains (at convergence, Rhat <= 1.05).
```

It yields a result of [79.8, 85.6].

Difference between the models

The separate model, in which the standard deviation and mean are considered as dependent on each of the existing groups yields results that are less realistic than the hierarchical one, in which the assumption that only a set of data represents the population is removed by establishing a multilevel approach. Finally, in the pooled model, the results are also better than in the separate one but not as good as in the hierarchical.