Week 6: Visualizing the Bayesian Workflow

Quynh Vu

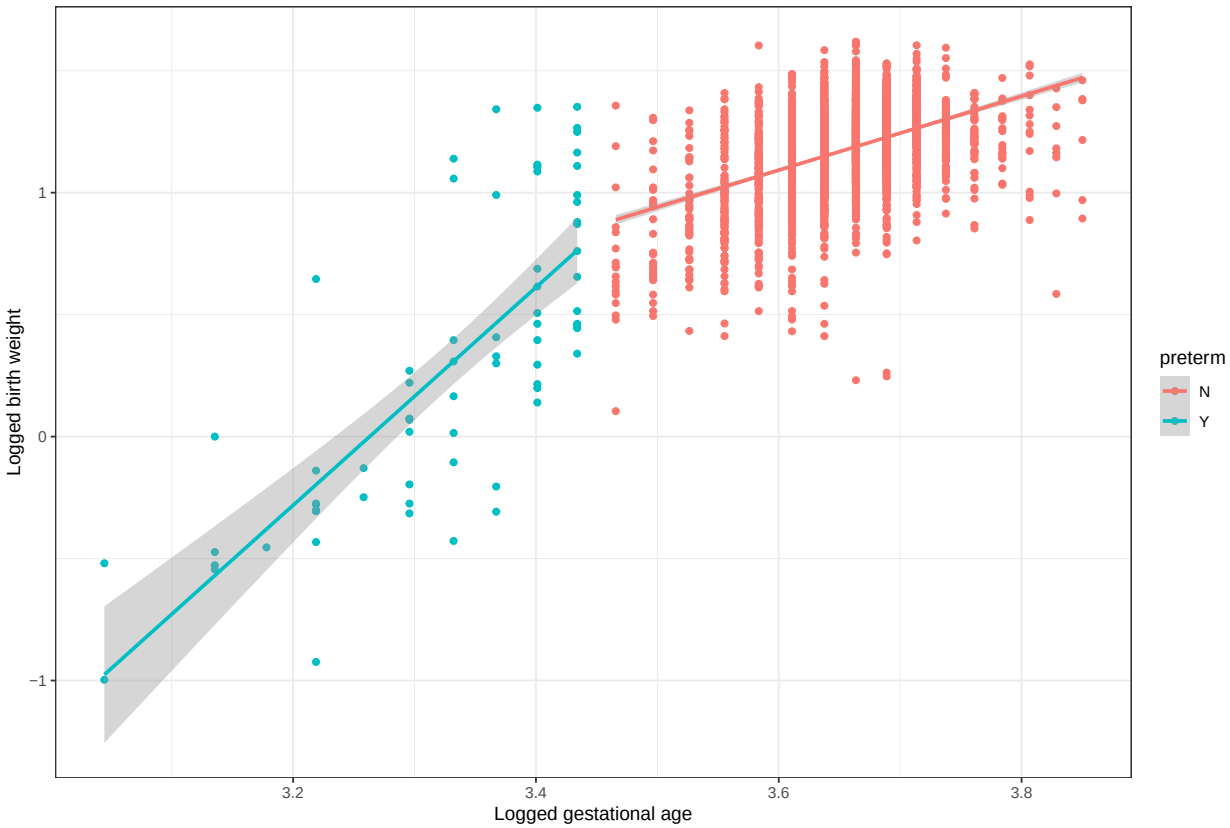
2023-02-24

Question 1: Use plots or tables to show three interesting observations about the data. Remember:

- Explain what your graph/ tables show
- Choose a graph type that's appropriate to the data type
- If you use `geom_smooth`, please also plot the underlying data

Feel free to replicate one of the scatter plots in the lectures as one of the interesting observations, as those form the basis of our models.

On the log scale, birth weight appears to increase as gestational age increases, and birth weight is higher after preterm. However, there are some exceptions that birth weight after preterm falls below the average birth weight during preterm and birth weight during preterm is higher than the average of birth weight after preterm.



The three interesting observations about the data are

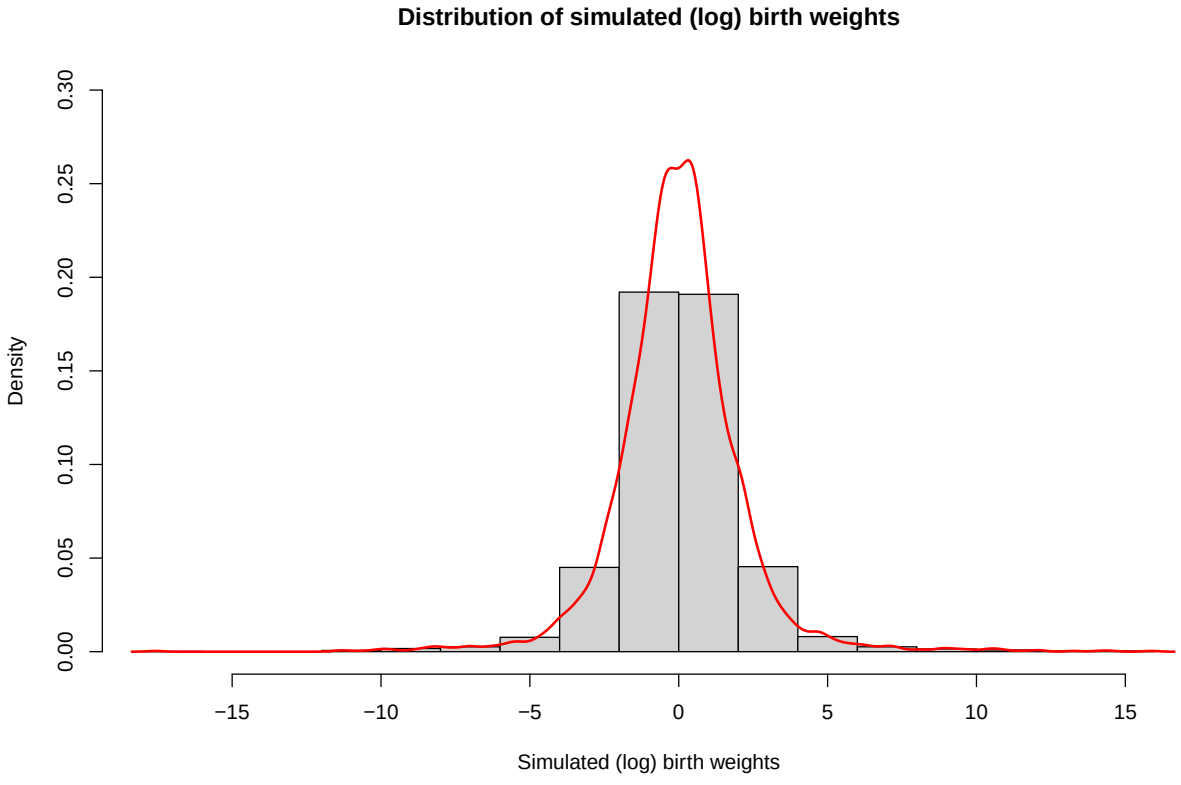
```
## # A tibble: 2 x 2
##   preterm avg_log_birthweight
##   <chr>                 <dbl>
## 1 N                     1.18
## 2 Y                     0.296
```

```
## # A tibble: 3 x 12
##   mager mracehisp meduc   bmi sex    gest birthw~1 ilive preterm log_b~2 log_g~3
##   <dbl>     <dbl> <dbl> <dbl> <chr> <dbl>    <dbl> <chr> <chr>     <dbl>   <dbl>
## 1    35         1     4  22.6 F        40     1.28 Y     N         0.247    3.69
## 2    39         2     5  23.4 M        29     3.83 Y     Y         1.34     3.37
## 3    34         2     3  37.9 M        40     1.3  Y     N         0.262    3.69
## # ... with 1 more variable: preterm_binary <dbl>, and abbreviated variable
## #   names 1: birthweight, 2: log_birthweight, 3: log_gest
```

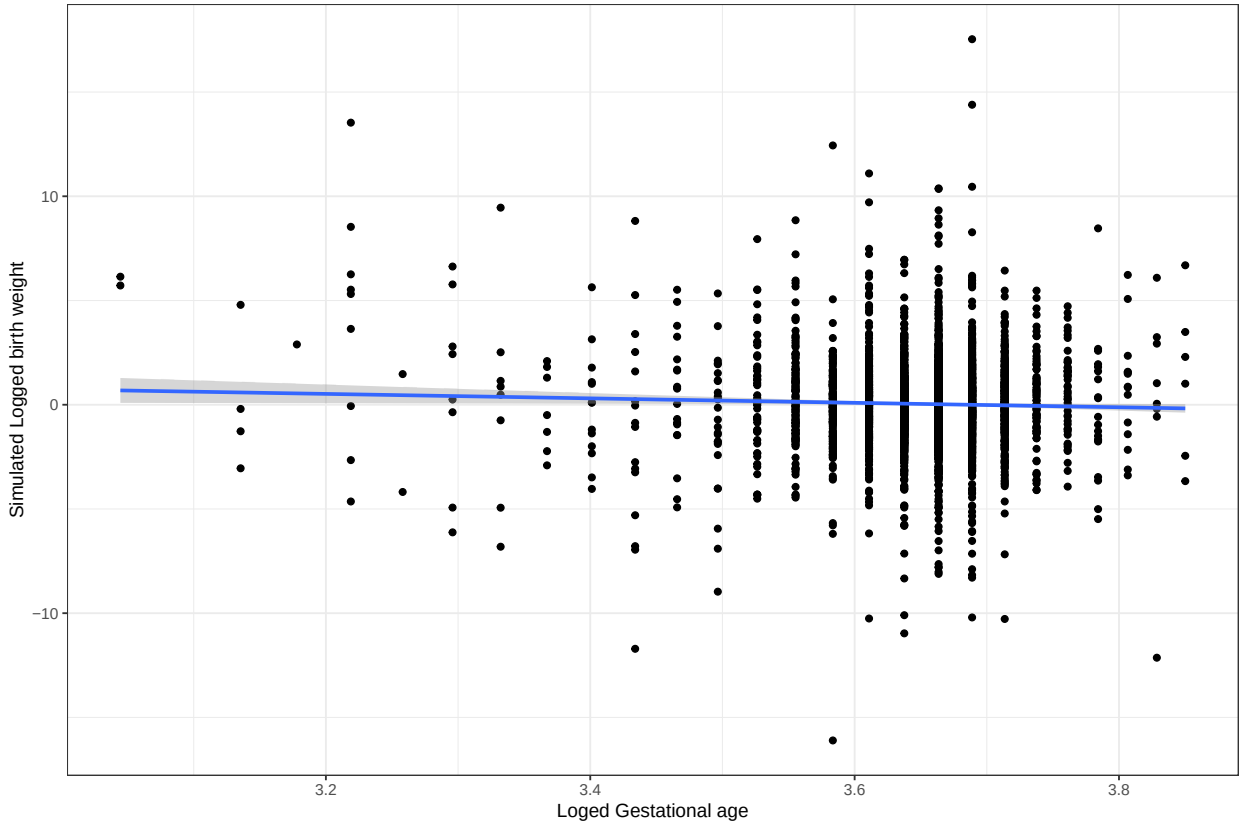
Question 2: Model 1 has log birth weight as a function of log gestational age $\log(y_i) \sim N(\beta_1 + \beta_2 \log(x_i), \sigma^2)$

Let's put some weakly informative priors on all parameters i.e. $\beta \sim N(0, 1)$ and $\sigma \sim N^+(0, 1)$ where the plus means positive values only i.e. Half Normal.

Plot the resulting distribution of simulated (log) birth weights



Plot ten simulations of (log) birthweights against gestational age



Run the model

```
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 0.000345 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 3.45 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
## Chain 1:
## Chain 1: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 1: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 1: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 1: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 1: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 1: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 1: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 1: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 1: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 1: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 1: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 1: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 0.633 seconds (Warm-up)
## Chain 1:                0.438 seconds (Sampling)
```

```
## Chain 1:                1.071 seconds (Total)
## Chain 1:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 0.000186 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 1.86 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 2: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 2: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 2: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 2: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 2: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 2: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 2: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 2: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 2: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 2: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 2: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 2:
## Chain 2:  Elapsed Time: 0.586 seconds (Warm-up)
## Chain 2:                0.491 seconds (Sampling)
## Chain 2:                1.077 seconds (Total)
## Chain 2:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 3).
## Chain 3:
## Chain 3: Gradient evaluation took 0.000179 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 1.79 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 3: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 3: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 3: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 3: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 3: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 3: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 3: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 3: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 3: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 3: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 3: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 0.523 seconds (Warm-up)
## Chain 3:                0.439 seconds (Sampling)
## Chain 3:                0.962 seconds (Total)
## Chain 3:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 4).
```

```
## Chain 4:
## Chain 4: Gradient evaluation took 0.00018 seconds
## Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 1.8 seconds.
## Chain 4: Adjust your expectations accordingly!
## Chain 4:
## Chain 4:
## Chain 4: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 4: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 4: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 4: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 4: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 4: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 4: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 4: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 4: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 4: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 4: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 4: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 4:
## Chain 4:  Elapsed Time: 0.492 seconds (Warm-up)
## Chain 4:                0.439 seconds (Sampling)
## Chain 4:                0.931 seconds (Total)
## Chain 4:
```

```
##              mean      se_mean          sd      2.5%       25%       50%
## beta[1] 1.1624783 8.160385e-05 0.002856578 1.1570200 1.1604786 1.1625011
## beta[2] 0.1437529 8.295075e-05 0.002912236 0.1381284 0.1416970 0.1436747
## sigma   0.1690330 1.113724e-04 0.001902828 0.1652694 0.1677842 0.1690763
##               75%     97.5%     n_eff      Rhat
## beta[1] 1.1644669 1.1681028 1225.3801 0.9978044
## beta[2] 0.1456716 0.1495180 1232.5721 0.9998714
## sigma   0.1702528 0.1727953  291.9066 1.0146111
```

Question 3: Based on model 1, give an estimate of the expected birthweight of a baby who was born at a gestational age of 37 weeks.

First, we center and standardized the log of gestational age of 37: $log(x_i) = \frac{log(37)-3.650894}{0.06723322} = -0.5945826$.

An estimate of the expected birthweight of a baby who was born at a gestational age of 37 weeks is $e^{1.1624783+0.1437529*(-0.5945826)} \approx 2.935874$. .which is very closed to the average observed birthweight of babies born at a gestational age of 37 weeks.

```
ds |> filter(gest == 37) |>
      summarise_at(vars(birthweight), list(expected_weight = mean))
```

```
## # A tibble: 1 x 1
##   expected_weight
##             <dbl>
## 1            3.05
```

Question 4: Write a stan model to run Model 2, and run it.

Model 2 has an interaction term between gestation and prematurity

$$\log(y_i) \sim N(\beta_1 + \beta_2 \log(x_i) + \beta_2 z_i + \beta_3 z_i + \beta_4 \log(x_i) z_i, \sigma^2)$$

```
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 0.000945 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 9.45 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
## Chain 1:
## Chain 1: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 1: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 1: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 1: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 1: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 1: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 1: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 1: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 1: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 1: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 1: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 1: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 3.954 seconds (Warm-up)
## Chain 1:                3.447 seconds (Sampling)
## Chain 1:                7.401 seconds (Total)
## Chain 1:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 0.000666 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 6.66 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 2: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 2: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 2: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 2: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 2: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 2: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 2: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 2: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 2: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 2: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 2: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 2:
## Chain 2:  Elapsed Time: 3.928 seconds (Warm-up)
## Chain 2:                3.386 seconds (Sampling)
## Chain 2:                7.314 seconds (Total)
## Chain 2:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 3).
```

```
## Chain 3:
## Chain 3: Gradient evaluation took 0.001671 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 16.71 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 3: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 3: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 3: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 3: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 3: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 3: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 3: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 3: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 3: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 3: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 3: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 3.986 seconds (Warm-up)
## Chain 3:                3.008 seconds (Sampling)
## Chain 3:                6.994 seconds (Total)
## Chain 3:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 4).
## Chain 4:
## Chain 4: Gradient evaluation took 0.000617 seconds
## Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 6.17 seconds.
## Chain 4: Adjust your expectations accordingly!
## Chain 4:
## Chain 4:
## Chain 4: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 4: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 4: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 4: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 4: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 4: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 4: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 4: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 4: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 4: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 4: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 4: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 4:
## Chain 4:  Elapsed Time: 3.804 seconds (Warm-up)
## Chain 4:                3.154 seconds (Sampling)
## Chain 4:                6.958 seconds (Total)
## Chain 4:
```

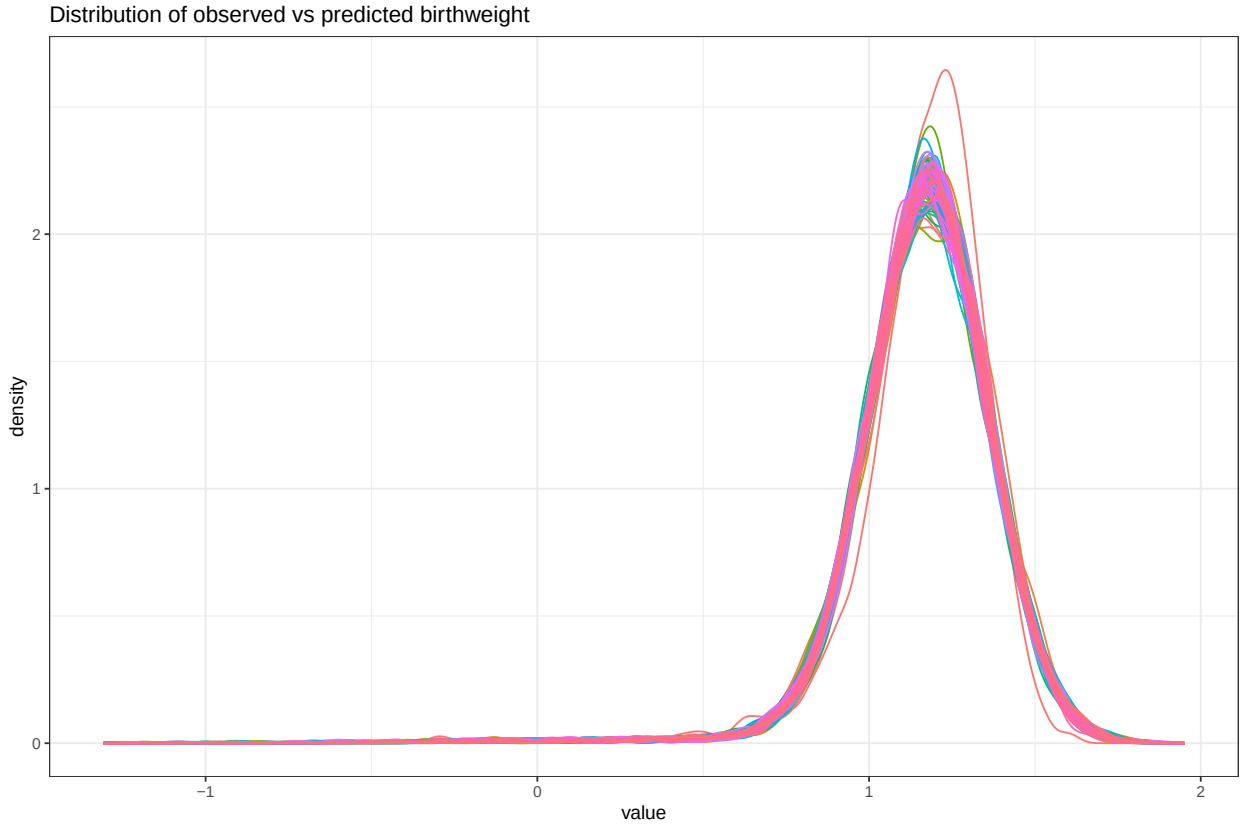
```
##              mean      se_mean          sd       2.5%        25%       50%
## beta[1] 1.1696329 8.021297e-05 0.002705139 1.16410383 1.16791913 1.1695478
## beta[2] 0.1018545 1.111662e-04 0.003424916 0.09508969 0.09961365 0.1019319
## beta[3] 0.5620695 3.406560e-03 0.062560942 0.43112646 0.52265217 0.5614275
## beta[4] 0.1982641 6.964438e-04 0.012807594 0.17144797 0.18979854 0.1986269
```

```
## sigma   0.1611971 8.785429e-05 0.001825790 0.15774991 0.15994557 0.1611909
##               75%     97.5%     n_eff     Rhat
## beta[1] 1.1714725 1.1748162 1137.3388 1.000638
## beta[2] 0.1040358 0.1087724  949.1923 1.002232
## beta[3] 0.6039584 0.6839901  337.2675 1.015352
## beta[4] 0.2062635 0.2232005  338.1917 1.013325
## sigma   0.1623667 0.1649513  431.8927 1.004553
```

Question 5 For reference I have uploaded some model 2 results. Check your results are similar.

```
##              mean      se_mean          sd       2.5%        25%       50%
## beta[1] 1.1697241 1.385590e-04 0.002742186 1.16453578 1.16767109 1.1699278
## beta[2] 0.5563133 5.835253e-03 0.058054991 0.43745504 0.51708255 0.5561553
## beta[3] 0.1020960 1.481816e-04 0.003669476 0.09459462 0.09997153 0.1020339
## beta[4] 0.1967671 1.129799e-03 0.012458398 0.17164533 0.18817091 0.1974114
## sigma   0.1610727 9.950037e-05 0.001782004 0.15784213 0.15978020 0.1610734
##               75%     97.5%     n_eff      Rhat
## beta[1] 1.1716235 1.1750167 391.67359 1.0115970
## beta[2] 0.5990427 0.6554967  98.98279 1.0088166
## beta[3] 0.1044230 0.1093843 613.22428 0.9978156
## beta[4] 0.2064079 0.2182454 121.59685 1.0056875
## sigma   0.1623019 0.1646189 320.75100 1.0104805
```

Question 6: Make a similar plot to the one above but for model 2, and not using the bayes plot in built function (i.e. do it yourself just with `geom_density`)



Question 7: Use a test statistic of the proportion of births under 2.5kg. Calculate the test statistic for the data, and the posterior predictive samples for both models, and plot the comparison (one plot per model).

The test statistics for the data are

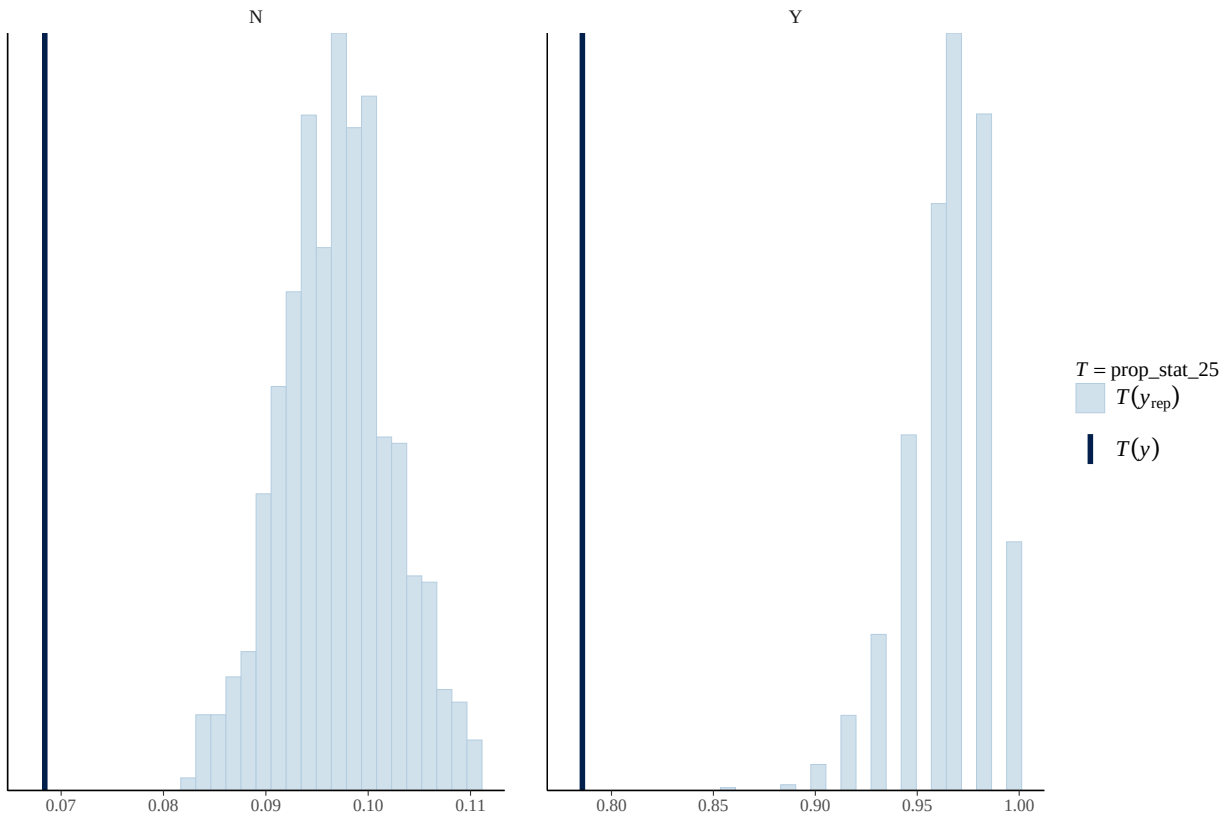
```
prop_stat_25(tab3$log_birthweight) # test statistic for group during preterm
```

```
## [1] 0.7857143
```

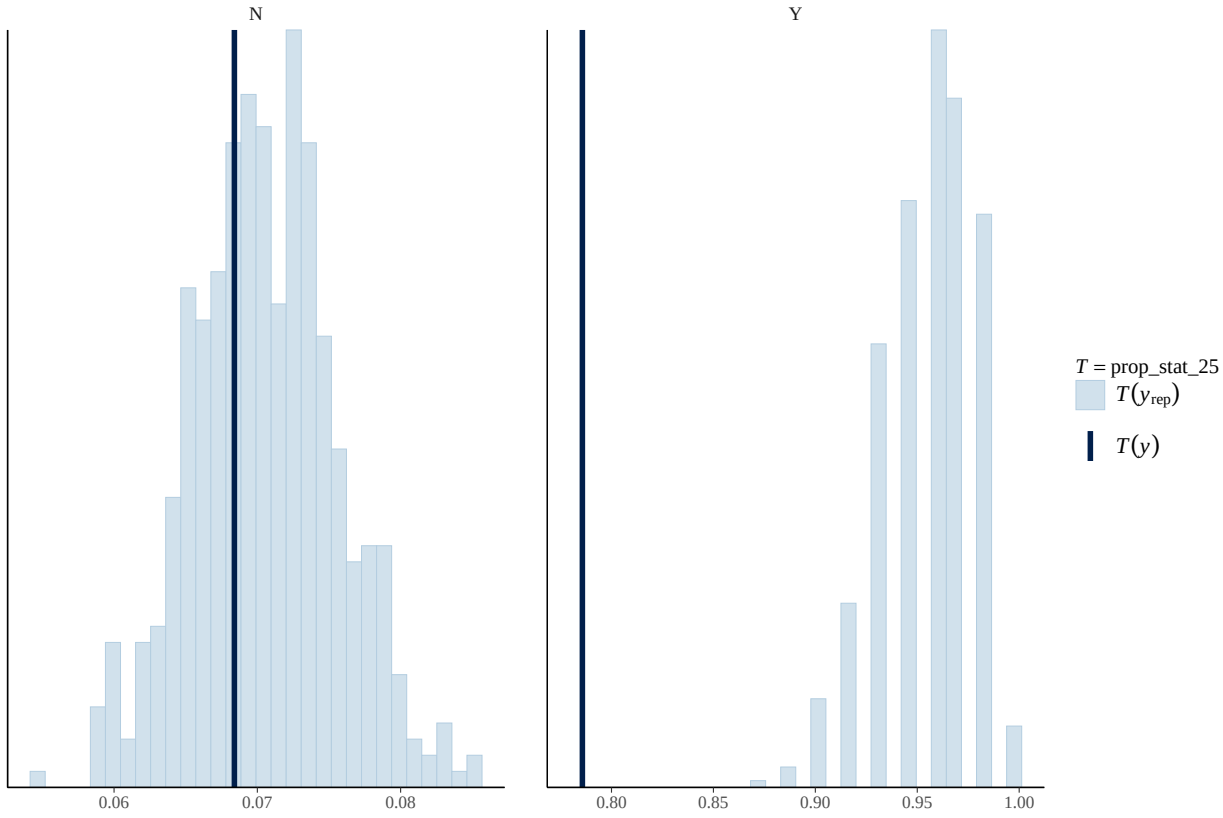
```
prop_stat_25(tab4$log_birthweight) # test statistic for group after preterm
```

```
## [1] 0.06839873
```

Model 1:


N
Y
0.07
0.08
0.09
0.10
0.11
0.80
0.85
0.90
0.95
1.00
T = prop_stat_25
T(y_rep)
T(y)

Model 2:


N
Y
0.06
0.07
0.08
0.80
0.85
0.90
0.95
1.00
T = prop_stat_25
T(y_rep)
T(y)

Question 8: Based on the original dataset, choose one (or more) additional covariates to add to the linear regression model. Run the model in Stan, and compare with Model 2 above on at least 2 posterior predictive checks.

```
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 0.000667 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 6.67 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
## Chain 1:
## Chain 1: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 1: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 1: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 1: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 1: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 1: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 1: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 1: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 1: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 1: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 1: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 1: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 3.916 seconds (Warm-up)
## Chain 1:                4.588 seconds (Sampling)
## Chain 1:                8.504 seconds (Total)
## Chain 1:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 0.000443 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 4.43 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 2: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 2: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 2: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 2: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 2: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 2: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 2: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 2: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 2: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 2: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 2: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 2:
## Chain 2:  Elapsed Time: 3.946 seconds (Warm-up)
## Chain 2:                4.366 seconds (Sampling)
## Chain 2:                8.312 seconds (Total)
## Chain 2:
```

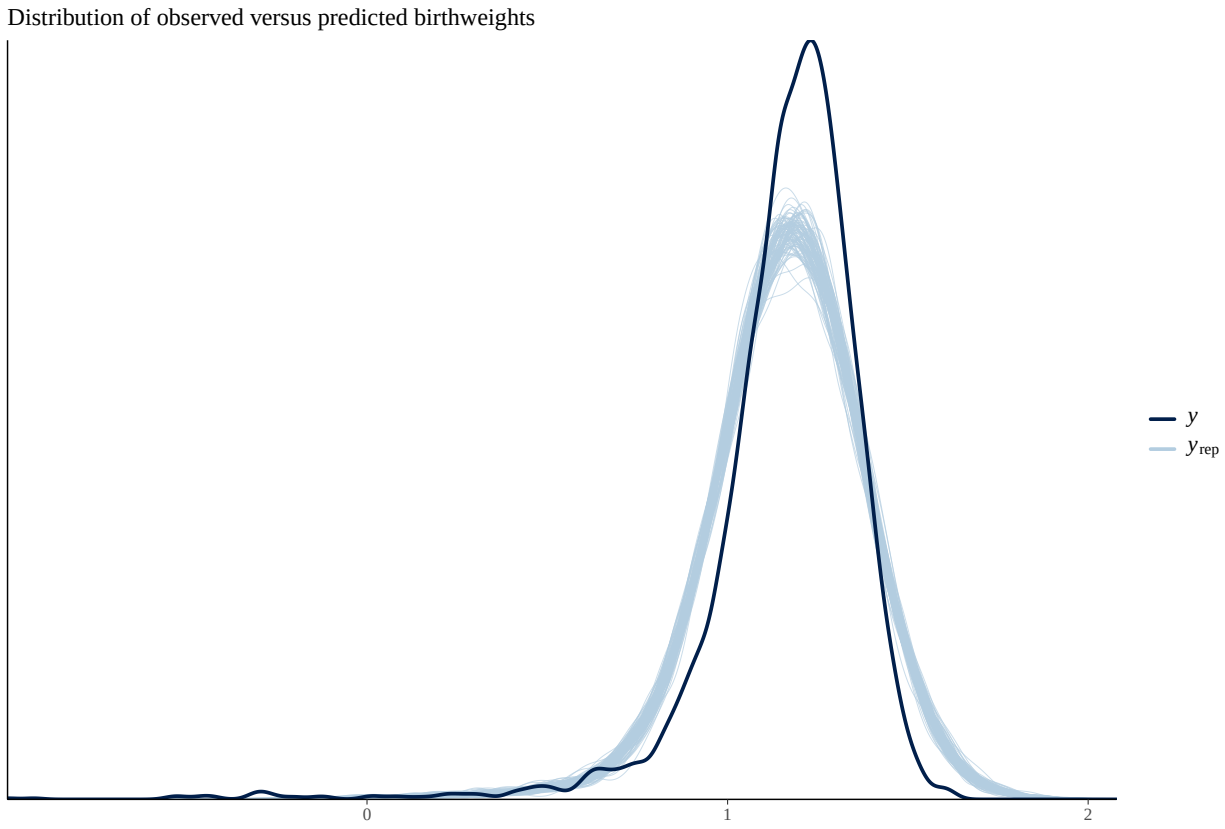
```
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 3).
## Chain 3:
## Chain 3: Gradient evaluation took 0.000408 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 4.08 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 3: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 3: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 3: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 3: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 3: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 3: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 3: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 3: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 3: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 3: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 3: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 4.504 seconds (Warm-up)
## Chain 3:                4.35 seconds (Sampling)
## Chain 3:                8.854 seconds (Total)
## Chain 3:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 4).
## Chain 4:
## Chain 4: Gradient evaluation took 0.000405 seconds
## Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 4.05 seconds.
## Chain 4: Adjust your expectations accordingly!
## Chain 4:
## Chain 4:
## Chain 4: Iteration:   1 / 500 [  0%]  (Warmup)
## Chain 4: Iteration:  50 / 500 [ 10%]  (Warmup)
## Chain 4: Iteration: 100 / 500 [ 20%]  (Warmup)
## Chain 4: Iteration: 150 / 500 [ 30%]  (Warmup)
## Chain 4: Iteration: 200 / 500 [ 40%]  (Warmup)
## Chain 4: Iteration: 250 / 500 [ 50%]  (Warmup)
## Chain 4: Iteration: 251 / 500 [ 50%]  (Sampling)
## Chain 4: Iteration: 300 / 500 [ 60%]  (Sampling)
## Chain 4: Iteration: 350 / 500 [ 70%]  (Sampling)
## Chain 4: Iteration: 400 / 500 [ 80%]  (Sampling)
## Chain 4: Iteration: 450 / 500 [ 90%]  (Sampling)
## Chain 4: Iteration: 500 / 500 [100%]  (Sampling)
## Chain 4:
## Chain 4:  Elapsed Time: 4.816 seconds (Warm-up)
## Chain 4:                5.463 seconds (Sampling)
## Chain 4:                10.279 seconds (Total)
## Chain 4:
##               mean      se_mean          sd      2.5%        25%        50%
## beta[1] 1.05116281 1.651825e-03 0.028599192 0.9976170 1.03143748 1.05100430
## beta[2] 0.14424307 9.841884e-05 0.002863328 0.1386688 0.14233937 0.14417253
```

```
## beta[3] 0.03366906 4.947107e-04 0.008574621 0.0157982 0.02794816 0.03365654
## sigma   0.16859683 9.836562e-05 0.001996482 0.1643057 0.16734764 0.16859194
##                75%      97.5%    n_eff     Rhat
## beta[1] 1.07023436 1.10975368 299.7640 1.014156
## beta[2] 0.14608401 0.14988261 846.4197 1.001482
## beta[3] 0.03973182 0.04927173 300.4189 1.014210
## sigma   0.16996889 0.17235255 411.9498 1.003098
```

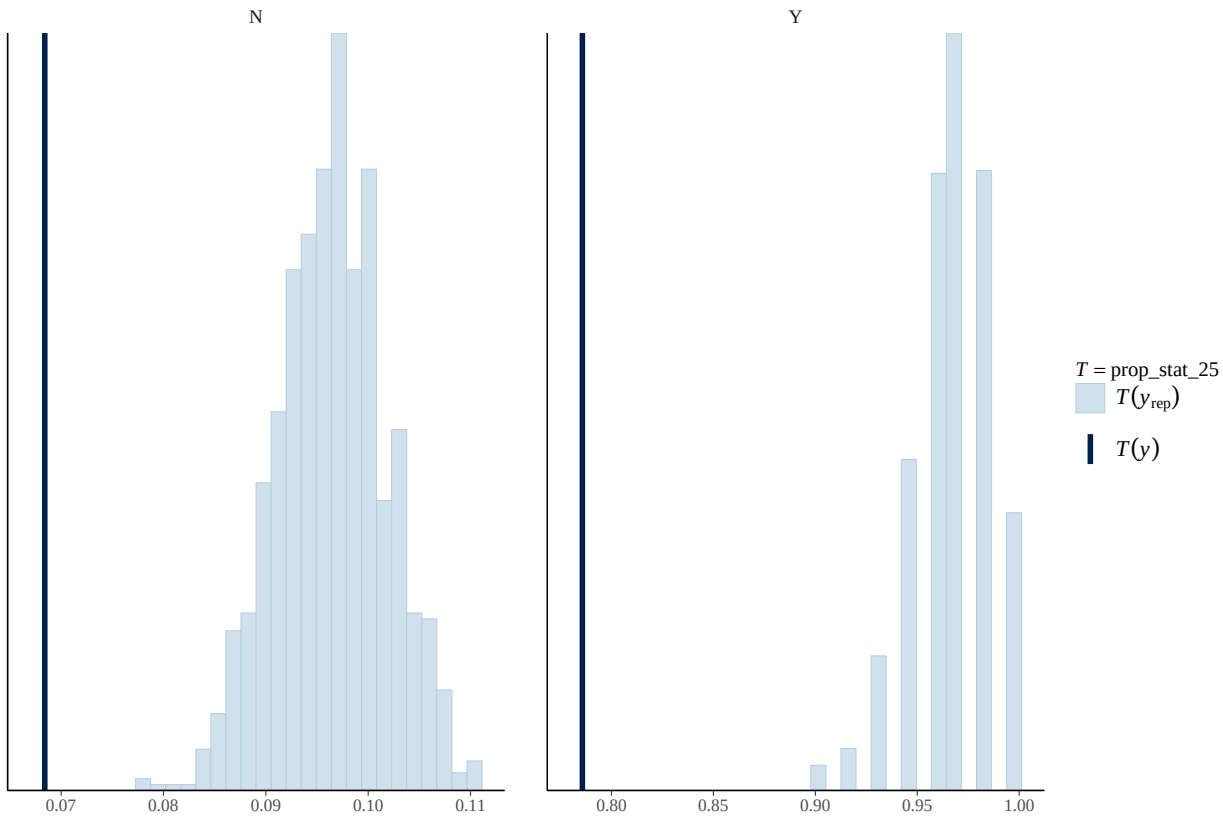
Posterior predictive checks:

By PPCs criterion, when comparing our original data set with 100 replicates, the new liner regression model with an additional covariate `log_bmi` does not fit the data as well as model 2 with an interaction term, as ones can see the density curves of those 100 replicates follow the observed density curve very loosely.

```
set.seed(1856)
yrep3 <- extract(mod3)[["log_weight_rep"]]
ppc_dens_overlay(y, yrep3[sample(nrow(yrep3), 100), ])  + ggtitle("Distribution of observed versus predi
```



When comparing the two models using test statistics of the proportion of births under 2.5kg, we still see that the new liner regression model with an additional covariate `log_bmi` does not fit the data as well as model 2 since the test statistics for the posterior predictive samples do not overlap or stay close to the test statistic for the data.


N
Y
0.07
0.08
0.09
0.10
0.11
0.80
0.85
0.90
0.95
1.00
T = prop_stat_25
T(y_rep)
T(y)