Posterior Summary

Jonathan Ma

2025-12-05

Table of Contents

1 Diagnostics

Model is loaded from an RDS object in the repository, so you dont have to wait a long time for it to run.

1.1 Convergence and Efficienct

```
##  Family: bernoulli
##   Links: mu = logit
## Formula: latestage ~ age + sex + raceth + grade + size_z + year_z + marry
+ (1 | patientid) + (1 | regionid)
##    Data: seer_df2 (Number of observations: 3031)
##   Draws: 4 chains, each with iter = 3000; warmup = 1500; thin = 1;
##          total post-warmup draws = 6000
##
## Multilevel Hyperparameters:
## ~patientid (Number of levels: 3026)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     9.11      1.85     5.69    13.01 1.01      494      593
##
## ~regionid (Number of levels: 3)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.79      0.89     0.02     3.23 1.00     2931     3694
##
## Regression Coefficients:
##                Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept         -4.32      1.36    -7.17    -1.75 1.00     1677     2623
## age01M04years     -0.34      1.89    -4.05     3.32 1.00     9072     5229
## age05M09years     -1.17      1.92    -4.82     2.58 1.00     9688     5284
## age10M14years      0.09      1.88    -3.47     3.91 1.00     8516     4916
## age15M19years      0.07      1.90    -3.65     3.87 1.00    10007     4664
## age20M24years     -1.28      1.83    -4.95     2.31 1.00     7836     4698
## age25M29years     -0.24      1.67    -3.52     3.03 1.00     6268     4781
## age30M34years      1.14      1.46    -1.81     4.08 1.00     4090     4215
## age35M39years      0.49      1.26    -2.05     2.92 1.00     3444     4116
## age40M44years      0.45      1.09    -1.69     2.63 1.00     3134     4143
## age45M49years      0.38      0.93    -1.44     2.19 1.00     2876     3588
## age50M54years      0.98      0.89    -0.73     2.76 1.00     2511     3412
## age55M59years     -0.12      0.84    -1.78     1.48 1.00     2555     3562
## age60M64years     -0.38      0.81    -1.99     1.16 1.00     2093     3913
## age65M69years      0.36      0.80    -1.22     1.93 1.00     2668     3376
## age70M74years     -0.23      0.82    -1.82     1.38 1.00     2305     3031
## age75M79years      1.42      0.86    -0.24     3.12 1.00     2719     3650
## age80M84years     -0.58      0.94    -2.49     1.24 1.00     2326     3739
## age85M89years     -0.05      1.04    -2.10     2.00 1.00     2787     3505
## age90Pyears       -0.62      1.18    -2.99     1.66 1.00     3452     4375
## sexMale            0.58      0.45    -0.28     1.51 1.00     2666     3249
## racethO            1.79      0.81     0.29     3.43 1.00     1990     3574
## racethW           -0.03      0.67    -1.37     1.25 1.00     2804     3382
```

```
## gradeStart         6.06      1.13     3.96     8.41 1.01      635      881
## size_z             1.22      0.30     0.69     1.87 1.01      793     1682
## year_z            -0.02      0.21    -0.44     0.39 1.00     2784     3506
## marryUnmarried     0.47      0.46    -0.40     1.41 1.00     2780     2985
##
## Draws were sampled using sample(hmc). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

We see no scale reduction (Rhat=1.0 in all cases) and ESS are all sufficiently large, and 1 divergent transition, but its small enough to be due to chance.

1.2 VIF

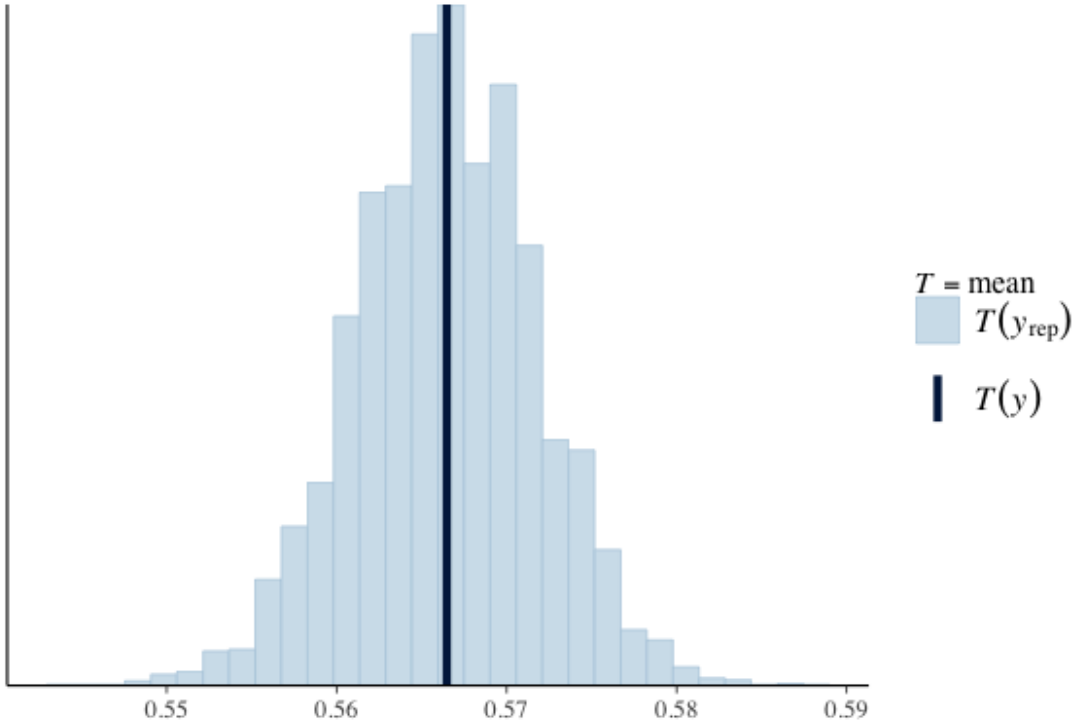
We check VIF, and confirm all values are reasonable, being around 1.

```
##            GVIF Df GVIF^(1/(2*Df))
## age    1.139013 19        1.003431
## sex    1.103208  1        1.050337
## raceth 1.058497  2        1.014314
## grade  1.053411  1        1.026358
## size_z 1.043417  1        1.021478
## year_z 1.006218  1        1.003104
## marry  1.093198  1        1.045561
```

1.3 Posterior Draws

Checking Posterior Draws

```
pp_check(breast, type = "stat")
```



The graph displays the posterior predictive distribution of the mean, comparing the observed mean of the outcome variable ($T(y)$, dark vertical line) to the distribution of means from replicated datasets generated under the fitted model ($T(y_{rep})$, light blue histogram). The close alignment between the observed and replicated means indicates that the model accurately captures the central tendency of the data. The symmetric and narrow distribution of $T(y_{rep})$ further suggests low variability and good calibration of the model's predictions.