730 Group Project

Rebekah Kristal with collaborators Amani Chehimi & Shane Fitzgerald

2024-11-30

Amani's Model: weighted linear regression

```
dat0 <- read_csv("FreqCategories.csv") %>%  mutate(Weight = Freq / sum(Freq))
```

```
## New names:
## Rows: 5462 Columns: 9
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (2): AgeCat, EduCat dbl (7): ...1, y, REGION, SEX, RACENEW, POORYN, Freq
## i Use ‘spec()‘ to retrieve the full column specification for this data. i
## Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## * ‘‘ -> ‘...1‘
```

```
fit <- brm(
  y | weights(Weight) ~ (1 | REGION + AgeCat + SEX + RACENEW + EduCat + POORYN),
  family = gaussian(),
  data = dat0,
  iter = 1000,
  chains = 4,
  cores = getOption("mc.cores", 4),
  seed = 12345
)
```

```
## Compiling Stan program...

## Start sampling

## Warning: There were 1 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
```

```
summary(fit)
```

```
## Warning: There were 1 divergent transitions after warmup. Increasing
## adapt_delta above 0.8 may help. See
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: y | weights(Weight) ~ (1 | REGION + AgeCat + SEX + RACENEW + EduCat + POORYN)
##    Data: dat0 (Number of observations: 5462)
##   Draws: 4 chains, each with iter = 1000; warmup = 500; thin = 1;
##          total post-warmup draws = 2000
##
## Multilevel Hyperparameters:
## ~AgeCat (Number of levels: 3)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     2.39      2.17     0.10     8.19 1.00     2064     1011
##
## ~EduCat (Number of levels: 4)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     2.45      2.24     0.11     8.14 1.00     2353     1366
##
## ~POORYN (Number of levels: 2)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     2.65      2.40     0.10     8.68 1.00     1563      867
##
## ~RACENEW (Number of levels: 6)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     2.59      2.37     0.09     8.59 1.00     2270     1027
##
## ~REGION (Number of levels: 4)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     2.40      2.22     0.06     7.96 1.00     1536      660
##
## ~SEX (Number of levels: 2)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     2.57      2.61     0.11     8.70 1.00     2284     1125
##
## Regression Coefficients:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     2.77      3.45    -4.11     9.44 1.00     2652     1345
##
## Further Distributional Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     6.05      4.70     1.92    16.92 1.01     2253     1161
##
## Draws were sampled using sampling(NUTS). 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).
```

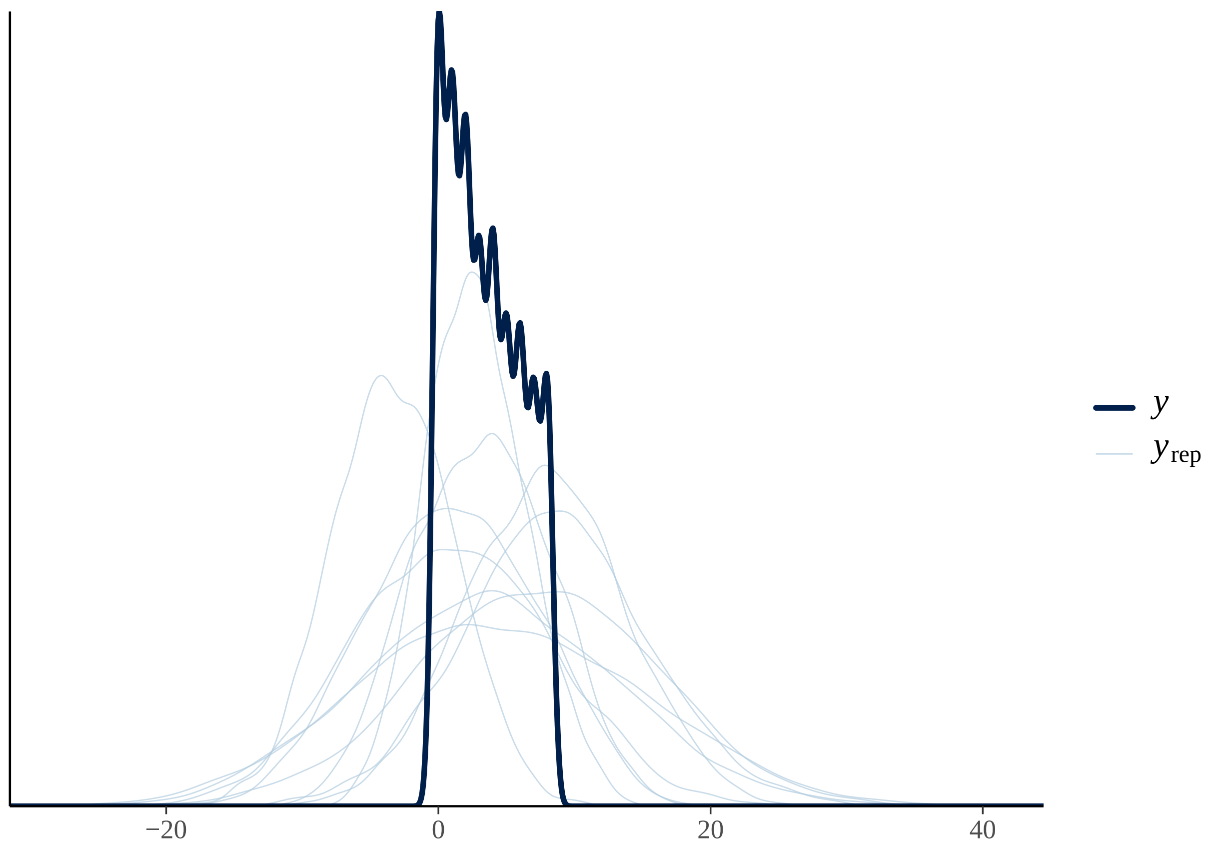
```
prior_summary(fit)
```

```
##                prior     class      coef   group resp dpar nlpar lb ub
##  student_t(3, 3, 3) Intercept
##  student_t(3, 0, 3)        sd                                    0
##  student_t(3, 0, 3)        sd             AgeCat                 0
##  student_t(3, 0, 3)        sd Intercept   AgeCat                 0
##  student_t(3, 0, 3)        sd             EduCat                 0
##  student_t(3, 0, 3)        sd Intercept   EduCat                 0
##  student_t(3, 0, 3)        sd             POORYN                 0
```

```
##  student_t(3, 0, 3)        sd Intercept  POORYN                  0
##  student_t(3, 0, 3)        sd           RACENEW                  0
##  student_t(3, 0, 3)        sd Intercept RACENEW                  0
##  student_t(3, 0, 3)        sd            REGION                  0
##  student_t(3, 0, 3)        sd Intercept  REGION                  0
##  student_t(3, 0, 3)        sd               SEX                  0
##  student_t(3, 0, 3)        sd Intercept     SEX                  0
##  student_t(3, 0, 3)     sigma                                    0
##        source
##       default
##       default
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##  (vectorized)
##       default
```

```r
pp_check(fit)
```

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```



Shane's models: weighted ordinal regression

```
newdata <- dat0
newdata<-mutate(newdata, weight.var=1/Freq) %>% mutate(REGION=as.factor(REGION)) %>% mutate(AgeCat=as.fa
```

```
#converting y's into factor variable, changing range from 0-8 to 1-9 to match with model output
newdata1<-mutate(newdata, y=y+1) %>% mutate(y, factor(y, ordered=TRUE))
mod3<-brm(y|weights(Freq)~REGION + AgeCat + SEX + RACENEW + EduCat + POORYN, data=newdata1, family=cumul
           chains = 4,
            iter = 2000, thin = 1)
```

```
## Compiling Stan program...
```

```
## Start sampling
```

```
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 0.007385 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 73.85 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
```

```
## Chain 1:
## Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 1: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 1: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 1: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 1: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 1: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 1: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 1: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 1: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 115.738 seconds (Warm-up)
## Chain 1:                108.486 seconds (Sampling)
## Chain 1:                224.224 seconds (Total)
## Chain 1:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 0.007122 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 71.22 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 2: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 2: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 2: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 2: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 2: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 2: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 2: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 2: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 2: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 2: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 2: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 2:
## Chain 2:  Elapsed Time: 105.166 seconds (Warm-up)
## Chain 2:                104.089 seconds (Sampling)
## Chain 2:                209.255 seconds (Total)
## Chain 2:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 3).
## Chain 3:
## Chain 3: Gradient evaluation took 0.006916 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 69.16 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 3: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 3: Iteration:  400 / 2000 [ 20%]  (Warmup)
```

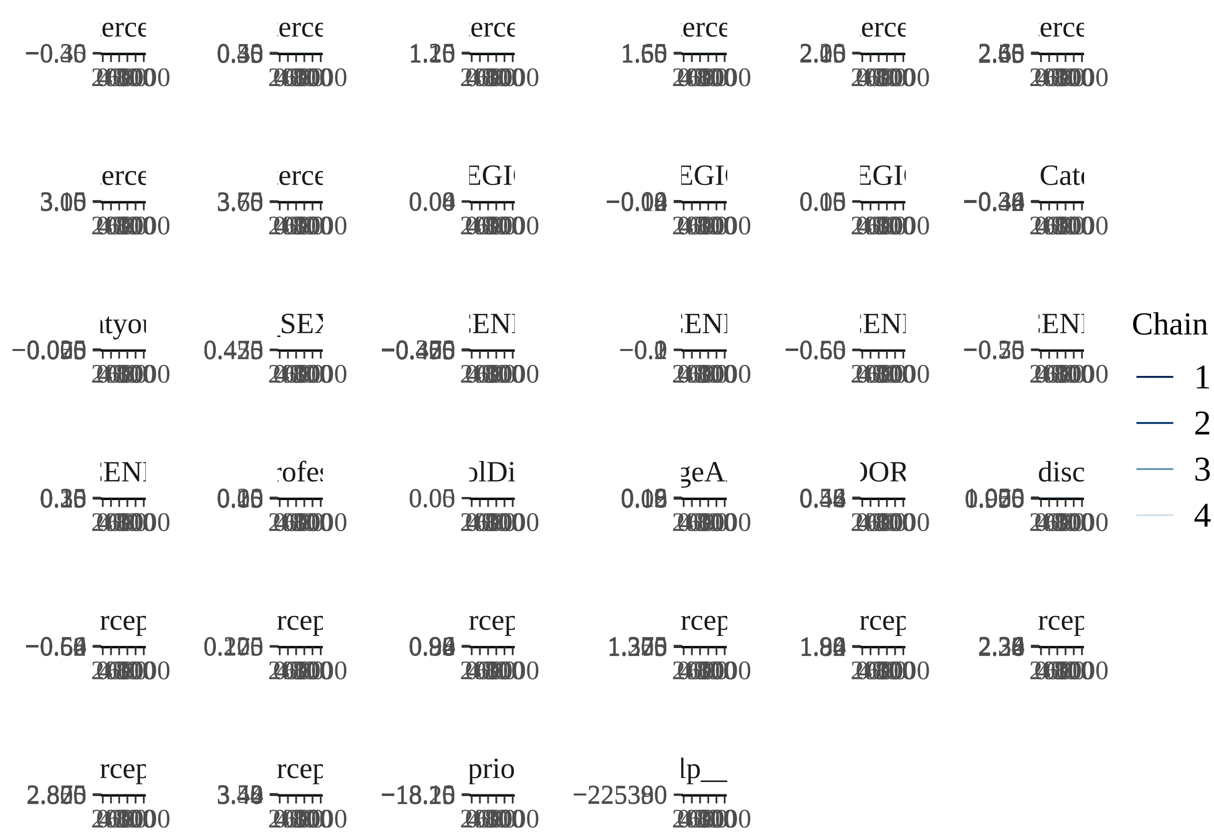
```
## Chain 3: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 3: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 3: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 3: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 3: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 3: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 3: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 3: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 3: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 114.983 seconds (Warm-up)
## Chain 3:                106.739 seconds (Sampling)
## Chain 3:                221.722 seconds (Total)
## Chain 3:
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 4).
## Chain 4:
## Chain 4: Gradient evaluation took 0.006565 seconds
## Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 65.65 seconds.
## Chain 4: Adjust your expectations accordingly!
## Chain 4:
## Chain 4:
## Chain 4: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 4: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 4: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 4: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 4: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 4: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 4: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 4: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 4: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 4: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 4: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 4: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 4:
## Chain 4:  Elapsed Time: 107.53 seconds (Warm-up)
## Chain 4:                108.189 seconds (Sampling)
## Chain 4:                215.719 seconds (Total)
## Chain 4:
```

```r
summary(mod3)
```

```
##  Family: cumulative
##   Links: mu = logit; disc = identity
## Formula: y | weights(Freq) ~ REGION + AgeCat + SEX + RACENEW + EduCat + POORYN
##    Data: newdata1 (Number of observations: 5462)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Regression Coefficients:
##                                    Estimate Est.Error l-95% CI u-95% CI
## Intercept[1]                          -0.35      0.02    -0.40    -0.31
## Intercept[2]                           0.43      0.02     0.39     0.47
## Intercept[3]                           1.15      0.02     1.11     1.20
```

```
## Intercept[4]                                1.56      0.02     1.52     1.61
## Intercept[5]                                2.10      0.02     2.06     2.15
## Intercept[6]                                2.54      0.02     2.50     2.59
## Intercept[7]                                3.07      0.02     3.02     3.12
## Intercept[8]                                3.68      0.03     3.63     3.74
## REGION2                                     0.04      0.02     0.00     0.07
## REGION3                                    -0.07      0.02    -0.10    -0.04
## REGION4                                     0.09      0.02     0.06     0.12
## AgeCatelderly                              -0.38      0.01    -0.41    -0.36
## AgeCatyoungAdult                           -0.03      0.01    -0.06    -0.00
## SEX2                                        0.45      0.01     0.43     0.47
## RACENEW200                                 -0.41      0.02    -0.44    -0.37
## RACENEW300                                 -0.10      0.05    -0.20    -0.01
## RACENEW400                                 -0.54      0.02    -0.59    -0.50
## RACENEW530                                 -0.53      0.11    -0.75    -0.32
## RACENEW541                                  0.26      0.04     0.19     0.33
## EduCatGraduateProfessionalorotherDegree     0.13      0.02     0.09     0.17
## EduCatHighschoolDiplomaGEDgraduate          0.02      0.02    -0.01     0.06
## EduCatSomecollegeAAorBachelorsDegree        0.12      0.02     0.09     0.16
## POORYN2                                     0.51      0.02     0.48     0.54
##                                         Rhat Bulk_ESS Tail_ESS
## Intercept[1]                            1.00     4044     2715
## Intercept[2]                            1.00     4005     2656
## Intercept[3]                            1.00     4073     2831
## Intercept[4]                            1.00     4116     2873
## Intercept[5]                            1.00     4267     2921
## Intercept[6]                            1.00     4420     3165
## Intercept[7]                            1.00     4371     3033
## Intercept[8]                            1.00     4476     2923
## REGION2                                 1.00     4463     3119
## REGION3                                 1.00     4252     3262
## REGION4                                 1.00     4049     3249
## AgeCatelderly                           1.00     7535     3390
## AgeCatyoungAdult                        1.00     8733     2907
## SEX2                                    1.00     7663     2695
## RACENEW200                              1.00     7005     2847
## RACENEW300                              1.00     6812     3046
## RACENEW400                              1.00     6316     2709
## RACENEW530                              1.00     7549     2755
## RACENEW541                              1.00     7633     3224
## EduCatGraduateProfessionalorotherDegree 1.00     3960     2975
## EduCatHighschoolDiplomaGEDgraduate      1.00     3774     2959
## EduCatSomecollegeAAorBachelorsDegree    1.00     3663     2845
## POORYN2                                 1.00     6597     3384
##
## Further Distributional Parameters:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## disc     1.00      0.00     1.00     1.00   NA       NA       NA
##
## Draws were sampled using sampling(NUTS). 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).
```

```r
mcmc_trace(mod3)
```



Check Model Assumptions for Amani's Model

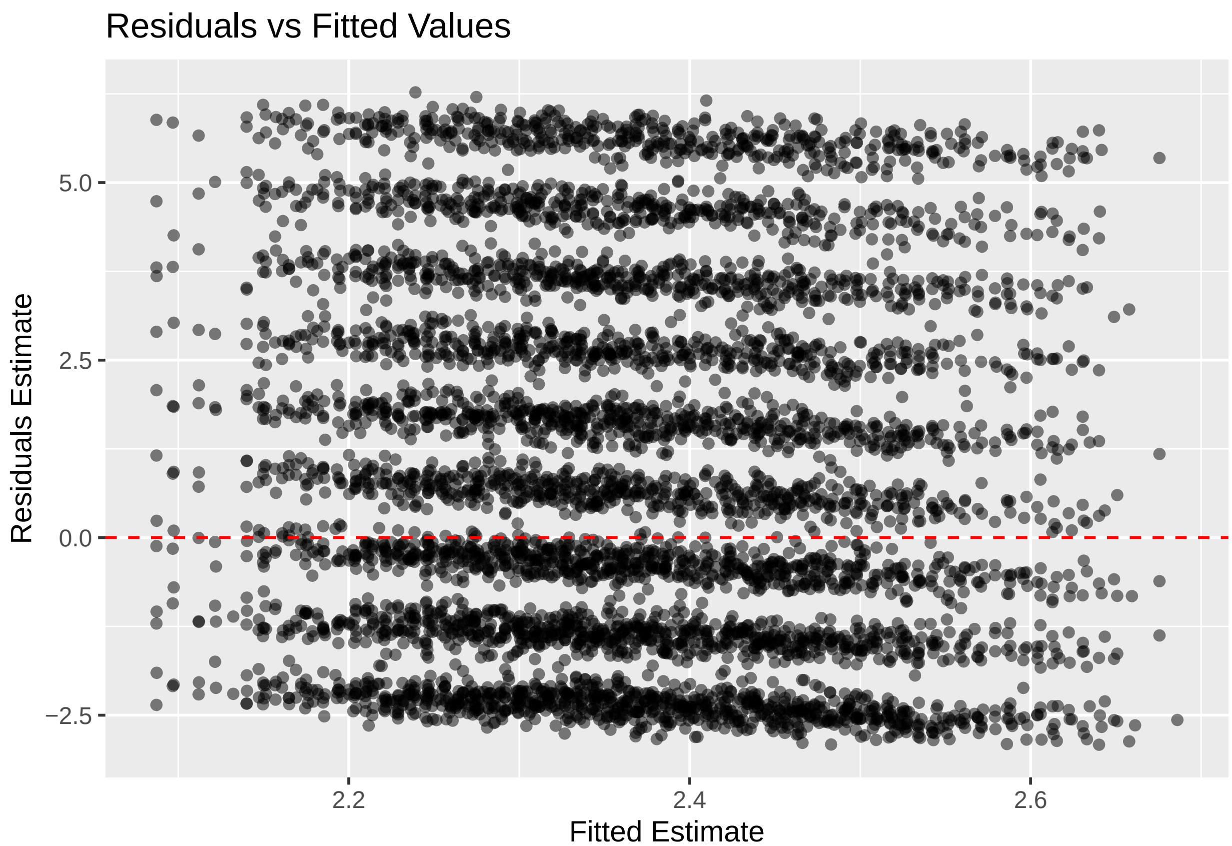
Residual Analysis

```r
residuals <- residuals(fit, type = "ordinary")
fitted_vals <- fitted(fit)

diagnostics_df <- data.frame(
  residuals = residuals,
  fitted = fitted_vals
)

colnames(diagnostics_df)
```

```
## [1] "residuals.Estimate"  "residuals.Est.Error" "residuals.Q2.5"
## [4] "residuals.Q97.5"     "fitted.Estimate"     "fitted.Est.Error"
## [7] "fitted.Q2.5"         "fitted.Q97.5"
```

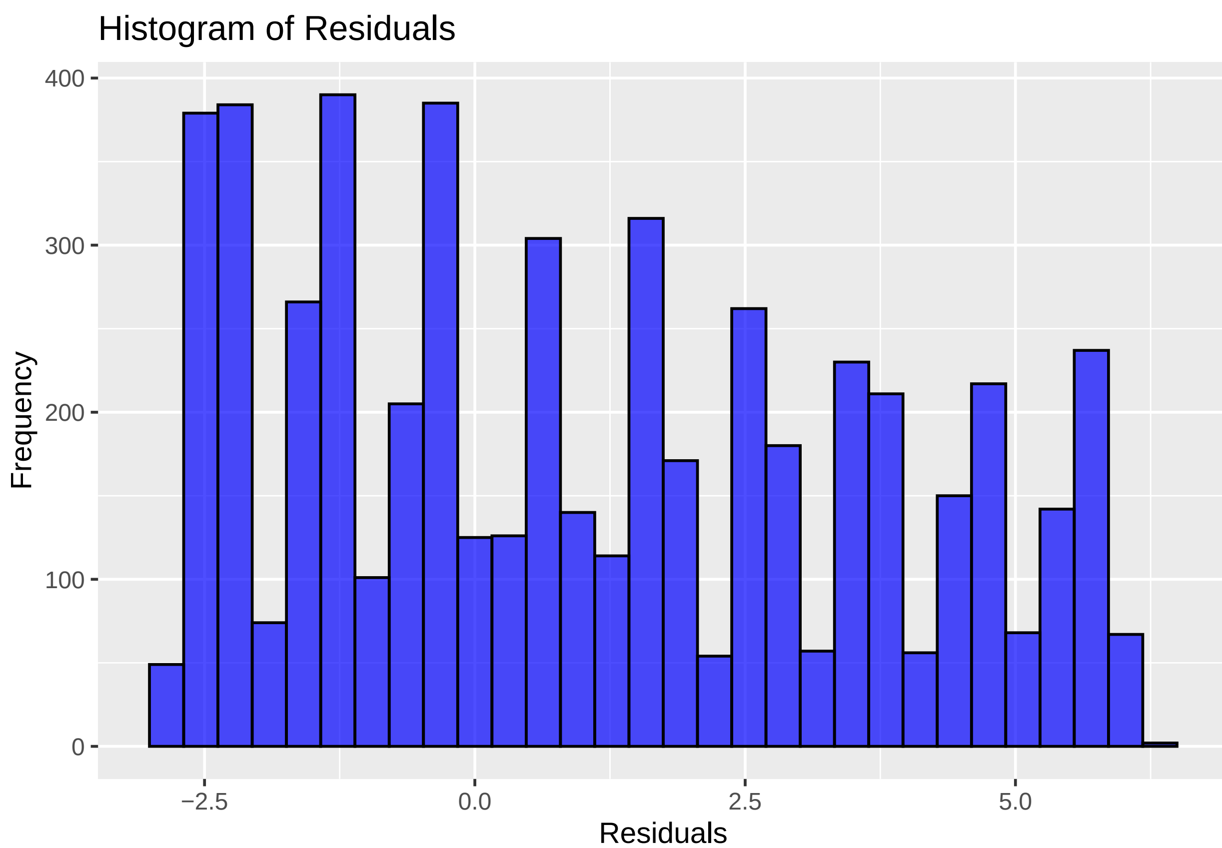
```r
# Residual vs. Fitted Plot
ggplot(data = diagnostics_df, aes(x = fitted.Estimate, y = residuals.Estimate)) +
  geom_point(alpha = 0.5) +
```

```r
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    title = "Residuals vs Fitted Values",
    x = "Fitted Estimate",
    y = "Residuals Estimate"
  )
```

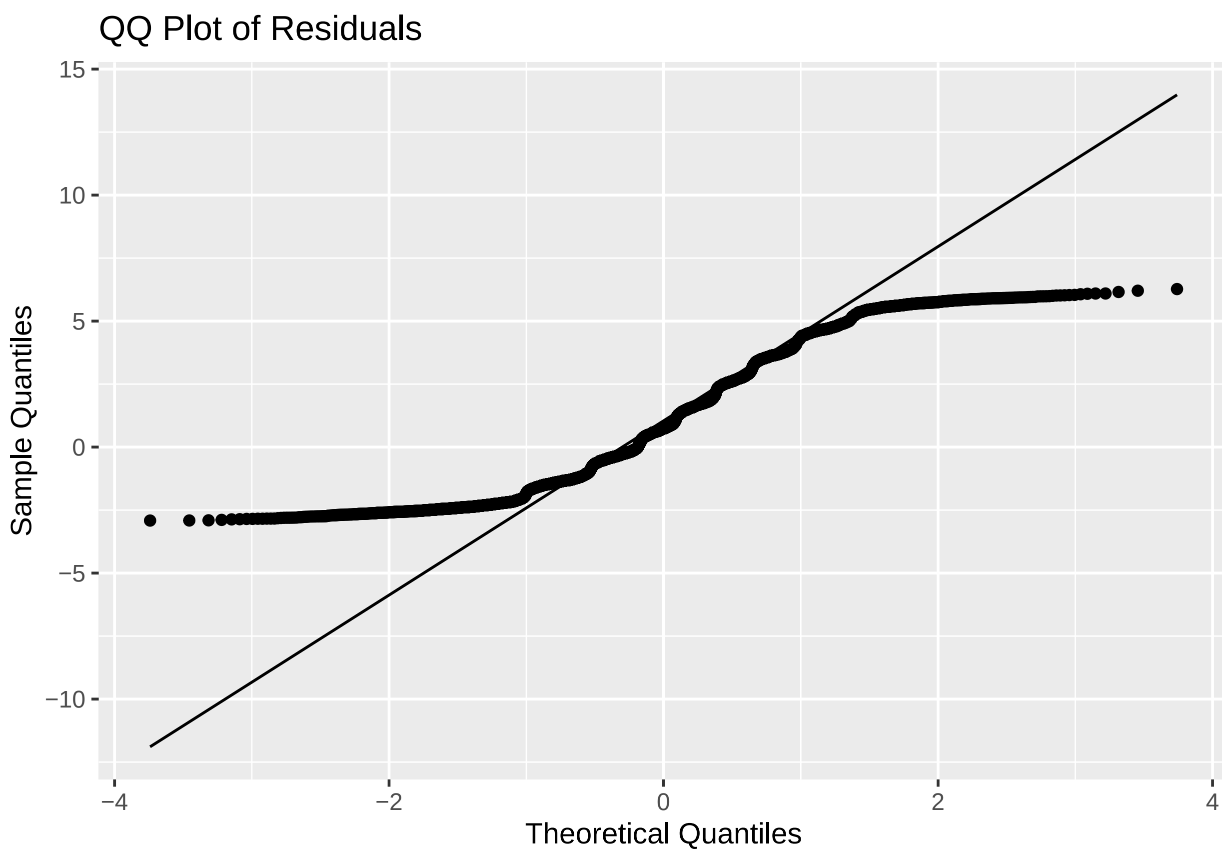


The "Residuals vs Fitted Values" plot reveals a potential issue with the model's assumptions. While the residuals are centered around the red dashed line at $y = 0$, indicating no systematic over- or under-prediction, the clear clustering or banding patterns suggest non-random residual distribution. This pattern might indicate heteroscedasticity (non-constant variance) or that the model is missing key variables or interactions needed to fully capture the underlying trends in the data. Additionally, the presence of distinct bands could point to the model's inability to account for certain systematic structures.

```r
#Histogram of the residuals
ggplot(data = data.frame(residuals = residuals), aes(x = residuals.Estimate)) +
  geom_histogram(bins = 30, color = "black", fill = "blue", alpha = 0.7) +
  labs(title = "Histogram of Residuals", x = "Residuals", y = "Frequency")
```




```r
# qqplot for residuals
ggplot(data = diagnostics_df, aes(sample = residuals.Estimate)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = "QQ Plot of Residuals", x = "Theoretical Quantiles", y = "Sample Quantiles")
```



The histogram of residuals shows a non-uniform distribution with visible peaks and valleys, further supporting the conclusion that the residuals do not follow a normal distribution. The clustering and uneven spread of residuals could point to underlying patterns in the data that the model has not fully captured, suggesting the need for further model refinement or consideration of alternative assumptions.

From the QQ plot, it is evident that the residuals deviate from the straight line at both ends, indicating that the residuals are not perfectly normally distributed. While the central portion of the residuals aligns fairly well with the theoretical quantiles, the deviations in the tails suggest potential issues with model assumptions, such as the presence of outliers or heavy-tailed distributions.

Fixed Effects and Coefficients

```r
fixef(fit) # Summary of Fixed Effects
```

```
##           Estimate Est.Error      Q2.5    Q97.5
## Intercept 2.773647  3.445892 -4.105874 9.437024
```

```r
posterior_summary(fit)
```

```
##                                                                  Estimate
## b_Intercept                                                   2.773647190
## sd_AgeCat__Intercept                                          2.394925035
## sd_EduCat__Intercept                                          2.453669299
## sd_POORYN__Intercept                                          2.648605917
## sd_RACENEW__Intercept                                         2.588552956
## sd_REGION__Intercept                                          2.396721969
```

```
## sd_SEX__Intercept                                                2.573721085
## sigma                                                            6.045196022
## Intercept                                                        2.773647190
## r_AgeCat[Adult,Intercept]                                       -0.100521171
## r_AgeCat[elderly,Intercept]                                     -0.164579076
## r_AgeCat[youngAdult,Intercept]                                  -0.082748418
## r_EduCat[Grade.12.or.less,.no.diploma,Intercept]                 0.014490005
## r_EduCat[Graduate,.Professional,.or.other.Degree,Intercept]     -0.080825320
## r_EduCat[High.school.Diploma,.GED,.graduate,Intercept]          -0.073734957
## r_EduCat[Some.college,.AA.or.Bachelor's.Degree,Intercept]       -0.133706579
## r_POORYN[1,Intercept]                                           -0.152415432
## r_POORYN[2,Intercept]                                            0.011720896
## r_RACENEW[100,Intercept]                                        -0.086194570
## r_RACENEW[200,Intercept]                                        -0.076592353
## r_RACENEW[300,Intercept]                                        -0.041140528
## r_RACENEW[400,Intercept]                                        -0.076072617
## r_RACENEW[530,Intercept]                                        -0.030660327
## r_RACENEW[541,Intercept]                                        -0.086065848
## r_REGION[1,Intercept]                                           -0.115921362
## r_REGION[2,Intercept]                                           -0.062962929
## r_REGION[3,Intercept]                                            0.008465024
## r_REGION[4,Intercept]                                           -0.018179146
## r_SEX[1,Intercept]                                              -0.008853820
## r_SEX[2,Intercept]                                              -0.033627545
## lprior                                                         -17.295516280
## lp__                                                           -45.678674687
##                                                                Est.Error
## b_Intercept                                                     3.445892
## sd_AgeCat__Intercept                                            2.170735
## sd_EduCat__Intercept                                            2.236577
## sd_POORYN__Intercept                                            2.400259
## sd_RACENEW__Intercept                                           2.370555
## sd_REGION__Intercept                                            2.221186
## sd_SEX__Intercept                                               2.606045
## sigma                                                           4.700051
## Intercept                                                       3.445892
## r_AgeCat[Adult,Intercept]                                       2.693323
## r_AgeCat[elderly,Intercept]                                     2.828556
## r_AgeCat[youngAdult,Intercept]                                  2.873745
## r_EduCat[Grade.12.or.less,.no.diploma,Intercept]                3.109126
## r_EduCat[Graduate,.Professional,.or.other.Degree,Intercept]     2.974180
## r_EduCat[High.school.Diploma,.GED,.graduate,Intercept]          3.065896
## r_EduCat[Some.college,.AA.or.Bachelor's.Degree,Intercept]       2.799065
## r_POORYN[1,Intercept]                                           2.898207
## r_POORYN[2,Intercept]                                           3.158893
## r_RACENEW[100,Intercept]                                        2.830314
## r_RACENEW[200,Intercept]                                        3.357839
## r_RACENEW[300,Intercept]                                        3.482794
## r_RACENEW[400,Intercept]                                        3.108606
## r_RACENEW[530,Intercept]                                        3.638183
## r_RACENEW[541,Intercept]                                        3.267318
## r_REGION[1,Intercept]                                           2.734135
## r_REGION[2,Intercept]                                           2.819009
## r_REGION[3,Intercept]                                           2.811103
```

```
## r_REGION[4,Intercept]                                          3.054715
## r_SEX[1,Intercept]                                             2.912173
## r_SEX[2,Intercept]                                             3.043892
## lprior                                                         2.635214
## lp__                                                           4.194502
##                                                                      Q2.5
## b_Intercept                                                   -4.10587378
## sd_AgeCat__Intercept                                           0.10371042
## sd_EduCat__Intercept                                           0.11487892
## sd_POORYN__Intercept                                           0.09695369
## sd_RACENEW__Intercept                                          0.08681715
## sd_REGION__Intercept                                           0.05597303
## sd_SEX__Intercept                                              0.11375415
## sigma                                                          1.92392723
## Intercept                                                     -4.10587378
## r_AgeCat[Adult,Intercept]                                     -6.07272177
## r_AgeCat[elderly,Intercept]                                   -6.39859294
## r_AgeCat[youngAdult,Intercept]                                -6.11342913
## r_EduCat[Grade.12.or.less,.no.diploma,Intercept]              -6.57252006
## r_EduCat[Graduate,.Professional,.or.other.Degree,Intercept]   -6.59410255
## r_EduCat[High.school.Diploma,.GED,.graduate,Intercept]        -6.75087937
## r_EduCat[Some.college,.AA.or.Bachelor's.Degree,Intercept]     -6.98558073
## r_POORYN[1,Intercept]                                         -6.52415193
## r_POORYN[2,Intercept]                                         -6.67628634
## r_RACENEW[100,Intercept]                                      -6.18849505
## r_RACENEW[200,Intercept]                                      -6.70482753
## r_RACENEW[300,Intercept]                                      -7.05905883
## r_RACENEW[400,Intercept]                                      -6.56214938
## r_RACENEW[530,Intercept]                                      -7.46160627
## r_RACENEW[541,Intercept]                                      -7.56958581
## r_REGION[1,Intercept]                                         -6.80164486
## r_REGION[2,Intercept]                                         -6.14031978
## r_REGION[3,Intercept]                                         -6.23923599
## r_REGION[4,Intercept]                                         -6.84265745
## r_SEX[1,Intercept]                                            -6.24479674
## r_SEX[2,Intercept]                                            -6.44681567
## lprior                                                       -24.03626549
## lp__                                                         -54.84127255
##                                                                   Q97.5
## b_Intercept                                                    9.437024
## sd_AgeCat__Intercept                                           8.189052
## sd_EduCat__Intercept                                           8.139685
## sd_POORYN__Intercept                                           8.681765
## sd_RACENEW__Intercept                                          8.591142
## sd_REGION__Intercept                                           7.961247
## sd_SEX__Intercept                                              8.702533
## sigma                                                         16.922253
## Intercept                                                      9.437024
## r_AgeCat[Adult,Intercept]                                      5.163678
## r_AgeCat[elderly,Intercept]                                    5.463970
## r_AgeCat[youngAdult,Intercept]                                 5.717069
## r_EduCat[Grade.12.or.less,.no.diploma,Intercept]               6.640484
## r_EduCat[Graduate,.Professional,.or.other.Degree,Intercept]    5.962604
## r_EduCat[High.school.Diploma,.GED,.graduate,Intercept]         6.558704
```

```
## r_EduCat[Some.college,.AA.or.Bachelor's.Degree,Intercept]     5.333408
## r_POORYN[1,Intercept]                                         5.642981
## r_POORYN[2,Intercept]                                         7.017319
## r_RACENEW[100,Intercept]                                      5.792705
## r_RACENEW[200,Intercept]                                      6.710483
## r_RACENEW[300,Intercept]                                      7.008554
## r_RACENEW[400,Intercept]                                      6.128344
## r_RACENEW[530,Intercept]                                      7.184433
## r_RACENEW[541,Intercept]                                      6.144274
## r_REGION[1,Intercept]                                         5.606092
## r_REGION[2,Intercept]                                         5.841721
## r_REGION[3,Intercept]                                         6.178506
## r_REGION[4,Intercept]                                         6.068296
## r_SEX[1,Intercept]                                            6.278923
## r_SEX[2,Intercept]                                            6.674882
## lprior                                                      -13.547874
## lp__                                                        -38.920884
```

Key Takeaways

Fixed Effects Interpretation
The intercept (2.63) shows high uncertainty, with a wide credible interval including zero. Group-level random effects show moderate variability, and residual variance ($\sigma = 5.99$) indicates significant unexplained variability.

Posterior Summary of Random Effects
Group-level deviations are small, often near zero, with credible intervals including zero, suggesting limited evidence for substantial effects.

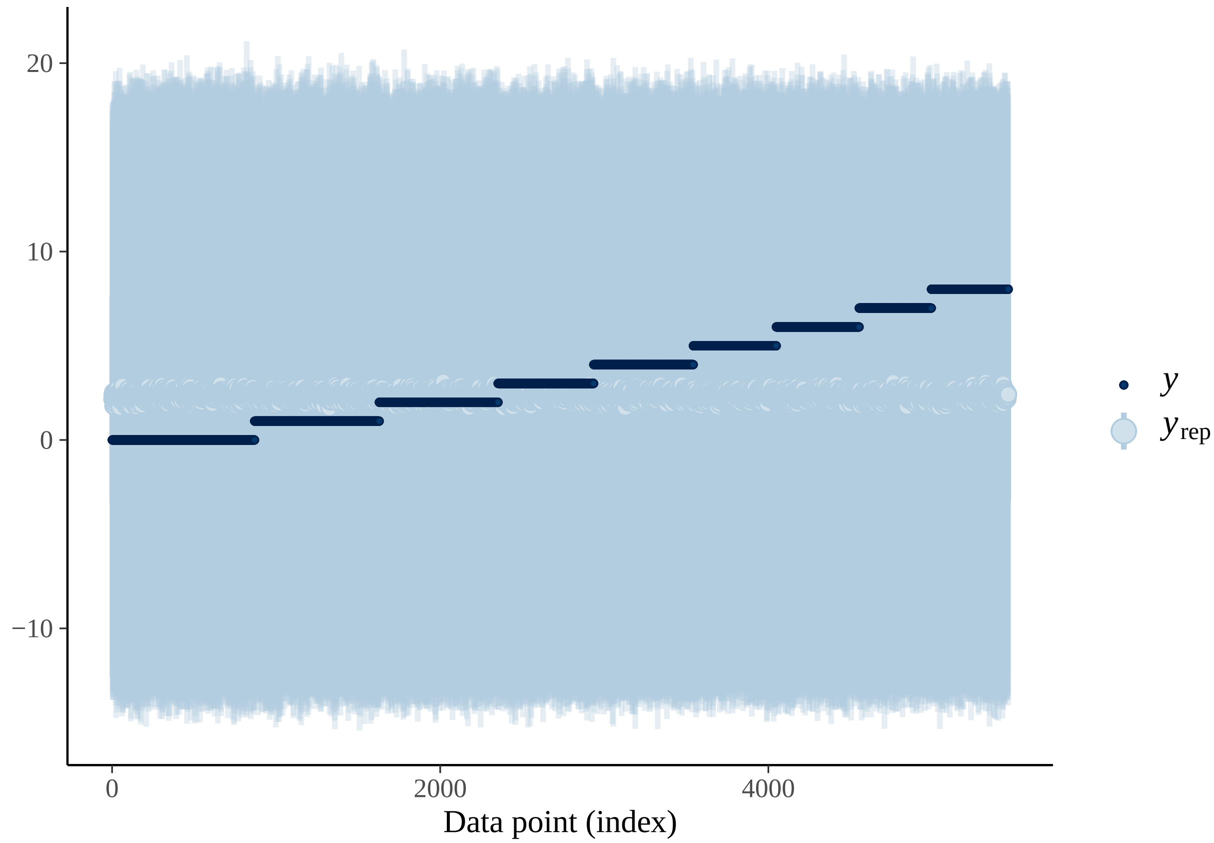
Model Fit and Variance
Unexplained variability remains high ($\sigma = 5.99$), with moderate group-level heterogeneity but wide credible intervals, highlighting the need for additional predictors.

PPC for Variable Combinations

```r
# PPC for Specific Covariate Combinations
pp_check(fit, type = "intervals", group = "AgeCat")
```

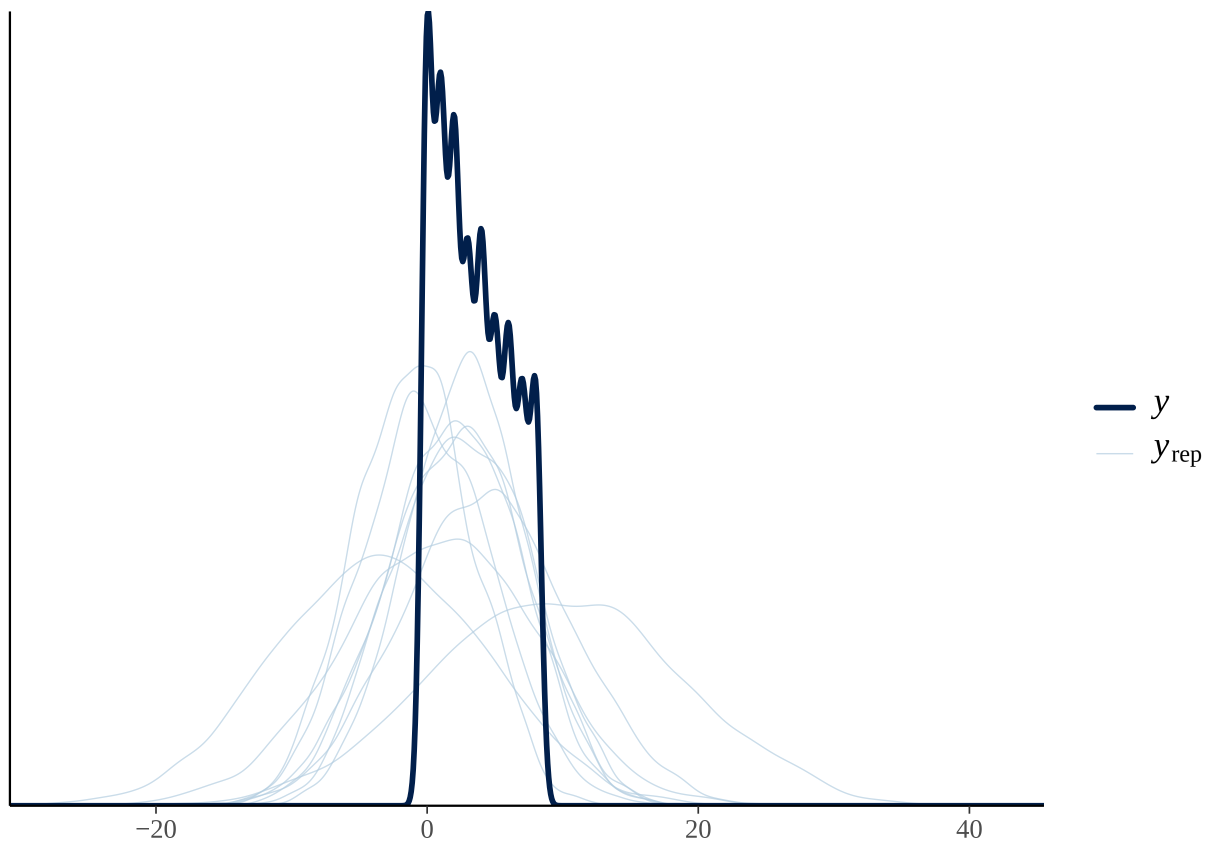
```
## Using all posterior draws for ppc type 'intervals' by default.
```

```
## Warning: The following arguments were unrecognized and ignored: group
```




```r
pp_check(fit, type = "dens_overlay")
```

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```



Key Takeaways

PPC Interval Plot Interpretation

The interval plot shows that the model generally captures the trends in the data, as most observed points (`y`) fall within the predicted intervals (`y_rep`). However, the stepwise pattern in the observed data, especially at higher indices, is not fully replicated by the model, indicating it may miss finer group-specific details.

PPC Density Overlay Plot Interpretation

The density plot highlights that the model captures the overall shape of the data but fails to replicate the sharp, stepwise peaks in the observed density (`y`). The predicted densities are smoother, and deviations are noticeable in the tails, suggesting the model struggles with abrupt changes and extreme values.

```r
# Export Fixed Effects Summary
fixed_effects_summary <- as.data.frame(fixef(fit))
write.csv(fixed_effects_summary, "fixed_effects_summary.csv", row.names = FALSE)

# Save Posterior Predictive Checks
ggsave("residuals_vs_fitted.png", width = 8, height = 6)
ggsave("histogram_residuals.png", width = 8, height = 6)
ggsave("ppc_intervals.png", width = 8, height = 6)
```

Analysis with Shane's Model

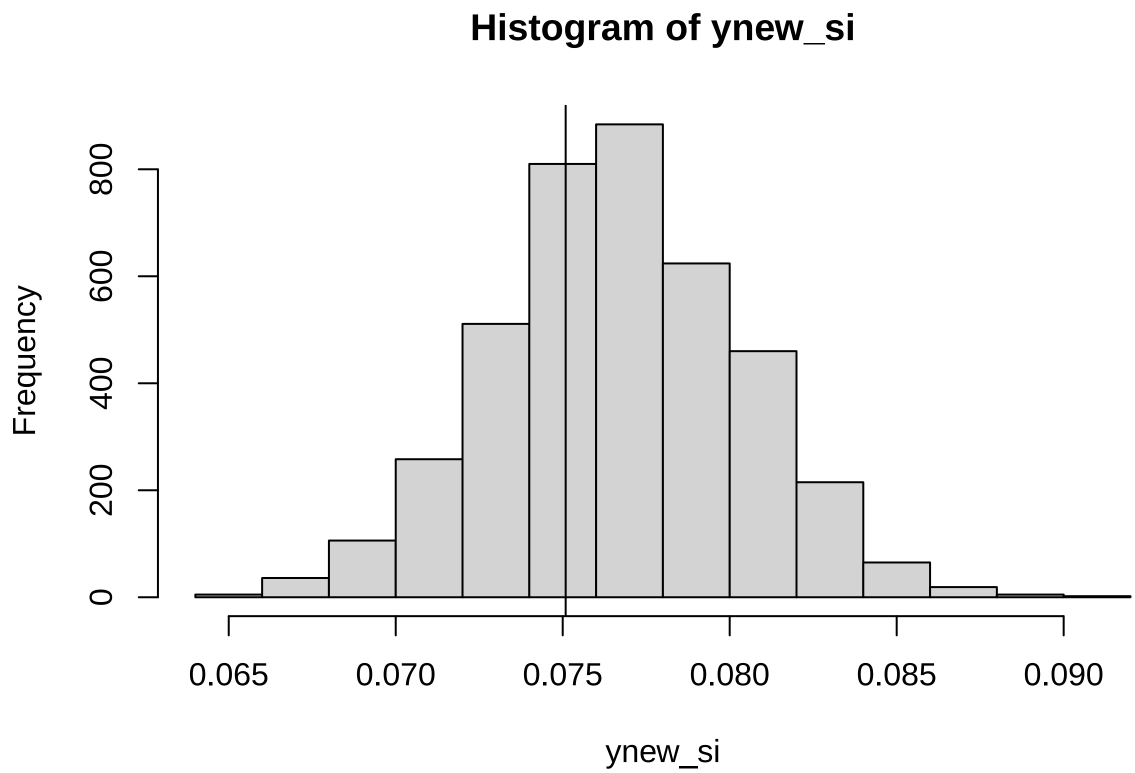
```r
#frequency table for observed data
observed_counts <- select(newdata1, c(y, Freq))
total_freq<-group_by(observed_counts, y) %>% summarise(total=sum(Freq))
observed_props<-mutate(total_freq, observed=total/sum(total)) %>% mutate(y=as.factor(y))



observed_props
```

```
## # A tibble: 9 x 3
##   y     total observed
##   <fct> <dbl>    <dbl>
## 1 1     45985   0.367
## 2 2     23279   0.186
## 3 3     20082   0.160
## 4 4      9250   0.0739
## 5 5      9403   0.0751
## 6 6      5513   0.0440
## 7 7      4483   0.0358
## 8 8      3187   0.0255
## 9 9      4043   0.0323
```

```r
#simulated datasets from model, creating function for summary statistic and getting summary statistic f

predicted_cats<-as.data.frame(posterior_predict(mod3))

get_sum_stat<-function(y, row){(sum(y==5))/nrow(row)}

tobs<-observed_props[5,3]
```

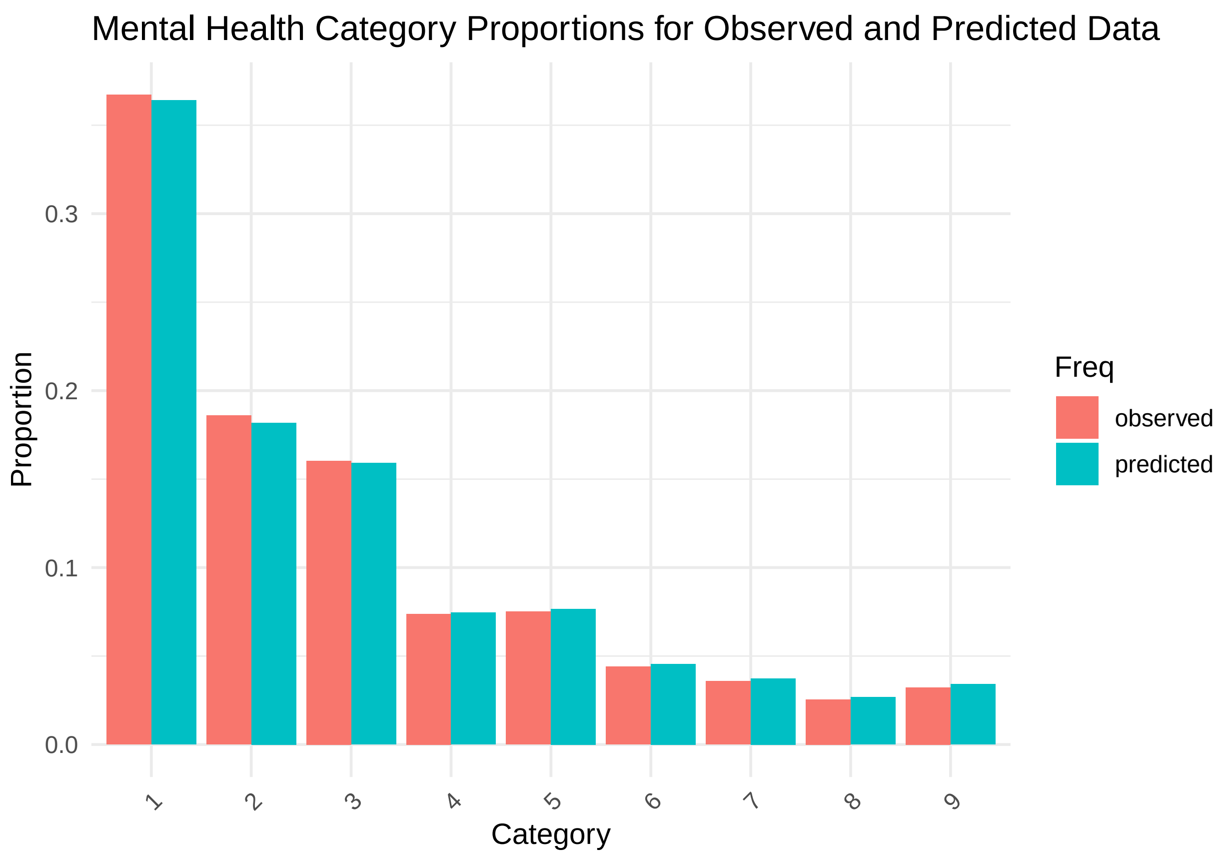
```r
#summary statistics for predicted datasets
ynew_si<-apply(predicted_cats, 1, get_sum_stat, newdata)
```

```r
#ppc for proportion of observations in category 5
hist(ynew_si)
abline(v = tobs)
```




```
#formatting for double barplot ggplot
posterior_preds_long <- predicted_cats %>%
  pivot_longer(cols = everything(), names_to = "chain", values_to = "predicted_category")

posterior_preds_long$predicted_category <- as.factor(posterior_preds_long$predicted_category)

category_counts <- table(posterior_preds_long$predicted_category)


category_counts_df <- as.data.frame(category_counts)
colnames(category_counts_df) <- c("y", "Count")

category_counts_prop<-mutate(category_counts_df, predicted=Count/(4000*5462))

combined<-left_join(observed_props, category_counts_prop, by="y")
combined1<-pivot_longer(combined, c(3,5), names_to = "Freq")
```

```
#plot of proportion of each category for observed and predicted data
ggplot(combined1, mapping=aes(x=y, y=value, fill=Freq))+
  geom_bar(stat="identity", position="dodge")+
  labs(title = "Mental Health Category Proportions for Observed and Predicted Data",
       x = "Category",
       y = "Proportion") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Mental Health Category Proportions for Observed and Predicted Data
Proportion
0.3
0.2
0.1
0.0
1
2
3
4
5
6
7
8
9
Category
Freq
observed
predicted