# HW2

Spencer Au

## Homework 2

## Data

The data for this homework assignment is *synthetic* data looking at different lifestyle and demographic variables and mental health/well-being for **500** people that are enrolled in **Company X's** wellness program. The following variables are included in the data:

- `anxiety`: subjective anxiety severity on a continuous scale of -10 to 10 (higher means more anxiety)

- `corgi`: average number of corgis pet per week.

- `group`: `A` for group A, `B` for group B; the "Joy Group" each person is assigned to. Joy Groups are social activity groups that meet once a week to do well-being related activities. Group `A` is led by Andrew, group `B` is led by Betty.

- `age` the age of the workers (in years)

- `hair_color`: hair color (`blonde`, `brown`, `black`, `other`)

- `veggies_week`: average servings of vegetables eaten per week.

- `therapy`: `0` if person does not currently attend therapy, `1` if they do

- `waived_insurance`:`0` if person did not waive their health insurance coverage, `1` if they did

- `sex`: sex assigned at birth

- `income_k`: household income in thousands of dollars.

- `married`: `married` if married, `divorced` if divorced, `single` if never married

- `mental_well`: mental well being, between -4 and 4, with higher scores indicating higher subjective mental well being.

# Analysis

## Question 1

- **Question 1**: (*Hypothesis Testing*) Andrew and Betty are arguing about who is better at leading their Joy Group. Joy groups were randomly assigned when employees joined the company. Is there evidence that Group A and Group B have the *same* mean anxiety score?

### Frequentist Analysis for Q1

```
# Frequentist Analysis
# null hypothesis: Group A and Group B have the same mean anxiety score
# alternative hypothesis: Group A and Group B have different mean anxiety scores

# t-test
q1_t_test <- t.test(health_data$anxiety ~ health_data$group)
```

Using a t-test, we find that $t = -0.5$ and the p-value is 0.6. Since the p-value is greater than 0.05, we fail to reject the null hypothesis that Group A and Group B have the same mean anxiety score. We then perform equivalence testing via a two one-sided t-test (TOST) to determine if the mean anxiety scores are "practically equivalent".

```
# TOST equiv test
# use with(data) to avoid having to use $ to access variables
q1_tost <- with(health_data, {
    tsum_TOST(
        m1 = mean(anxiety[group == "A"]),
        sd1 = sd(anxiety[group == "A"]),
        n1 = sum(group == "A"),
        m2 = mean(anxiety[group == "B"]),
        sd2 = sd(anxiety[group == "B"]),
        n2 = sum(group == "B"),
        low_eqbound = -0.1 * sd(anxiety),
        high_eqbound = 0.1 * sd(anxiety)
    )
})
```

**Bayesian Analysis for Q1**

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-5.7873 -2.2943 -0.5903 -0.3830  1.6770  4.6868


    Min.  1st Qu.   Median    Mean  3rd Qu.     Max.
-11.6076  -4.9243   0.1643  -0.1931   4.3572  10.2675
```

```
[1] "Standard Dev for A: 2.38"
```

```
[1] "Standard Dev for B: 5.72"
```

```
[1] "Overall Standard Dev: 4.32"
```

For Bayesian Hypothesis Testing, we use two models: one assumes a treatment effect (difference in mean anxiety scores between Group A and Group B), while the other assumes no treatment effect (null hypothesis). Based on the EDA, we set a prior of normal(0, 4) for the intercept, reflecting the overall standard deviation of 4.32; normal(0, 3) for the group effect, as it balances the differing standard deviations of Groups A and B; and student_t(3, 0, 4) for residual noise, with 3 degrees of freedom to account for outliers and the observed variability.

```
# model with treatment effect
invisible(q1_fit1_alt <- brm(
    formula = anxiety ~ group,
    data = health_data,
    family = gaussian(),
    prior = q1_priors,
    # update this since save_all_pars from slides is deprecated
    save_pars = save_pars(all = TRUE),
    silent = 2, refresh = 0
))
```

```
Trying to compile a simple C file


Running /Library/Frameworks/R.framework/Resources/bin/R CMD SHLIB foo.c
using C compiler: 'Apple clang version 16.0.0 (clang-1600.0.26.4)'
using SDK: 'MacOSX15.1.sdk'
clang -arch arm64 -I"/Library/Frameworks/R.framework/Resources/include" -DNDEBUG   -I"/Librar
In file included from <built-in>:1:
In file included from /Users/spencerau/Library/R/arm64/4.4/library/StanHeaders/include/stan/r
```

```
In file included from /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/R
In file included from /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/R
/Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/RcppEigen/include/Eigen/
  679 | #include <cmath>
      |          ^~~~~~~
1 error generated.
make: *** [foo.o] Error 1
```

```r
# model with no treatment effect
invisible(q1_fit2_null <- brm(
    formula = anxiety ~ 1,
    data = health_data,
    family = gaussian(),
    prior = c(
        prior(normal(0, 4), class = "Intercept"),
        prior(student_t(3, 0, 4), class = "sigma")
    ),
    save_pars = save_pars(all = TRUE),
    silent = 2, refresh = 0
))
```

```
Trying to compile a simple C file


Running /Library/Frameworks/R.framework/Resources/bin/R CMD SHLIB foo.c
using C compiler: 'Apple clang version 16.0.0 (clang-1600.0.26.4)'
using SDK: 'MacOSX15.1.sdk'
clang -arch arm64 -I"/Library/Frameworks/R.framework/Resources/include" -DNDEBUG    -I"/Libra
In file included from <built-in>:1:
In file included from /Users/spencerau/Library/R/arm64/4.4/library/StanHeaders/include/stan/r
In file included from /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/R
In file included from /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/R
/Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/RcppEigen/include/Eigen/
  679 | #include <cmath>
      |          ^~~~~~~
1 error generated.
make: *** [foo.o] Error 1
```

We then compare the two models, calculating their posterior probabilities assuming each model has an equal probability (50%) of being true. We also calculate the Bayes Factor to determine which model is more likely.

## Question 2

- **Question 2**: (*Parameter Estimation*) *assuming* the DAG shown below is correct, and using *covariates* in a regression (GLM) model to adjust for variables, what is the **direct causal effect** of **petting corgis** on **mental well-being**? Be sure to include both a point estimate and interval estimate.
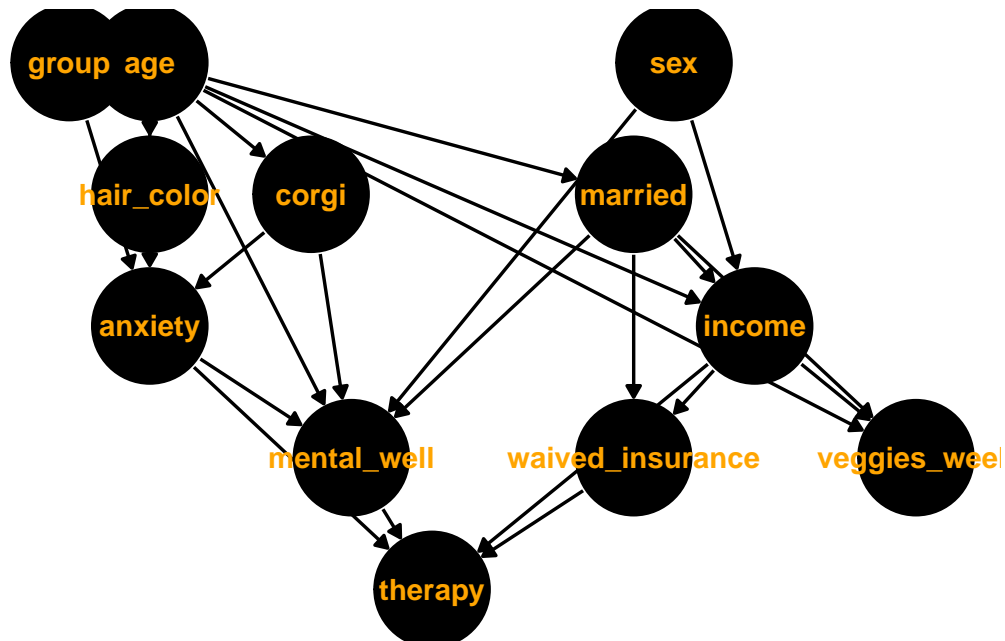


### Frequentist Analysis for Q2

For the Frequentist analysis, we use a Generalized Linear Model (GLM) to estimate the direct causal effect of petting corgis on mental well-being. We control for age due to a fork backdoor path and anxiety due to a chain backdoor path. We do not need to control for income and marriage status as controlling for age already blocks this path.

```
# control for age due to fork backdoor
# control for anxiety due to chain backdoor
# Income - Adjusting for age already blocks this path
# Married - controlling for age already blocks this path.
q2_glm <- glm(mental_well ~ corgi + age + anxiety,
              data = health_data,
              family = gaussian())
```

**Bayesian Analysis for Q2**

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -3.534   3.853   5.754   5.905   8.229  15.465
```

```
[1] "Standard Dev for Mental Wellness: 3.27"
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.000   2.294   3.000  12.000
```

```
[1] "Standard Dev for Corgi Pets: 1.68"
```

For the Bayesian analysis, we use a Bayesian Regression Model (BRM) to estimate the direct causal effect of petting corgis on mental well-being. We control for the same variables as in the Frequentist analysis: age and anxiety. We set the priors to normal(6, 3) for the intercept, reflecting the baseline mental well being mean of 5.9 with a standard deviation of 3.3, normal(1, 2) to account for an assumption that corgi interaction will have a positive effect on mental well-being with some variability (2) for effectiveness, and student_t(3, 0, 3) for residual noise, with 3 degrees of freedom to account for outliers and accounting for the observed standard deviation of 3.3 for mental well-being.

```
q2_bayes <- brm(mental_well ~ corgi + age + anxiety,
                data = health_data,
                family = gaussian(),
                prior = q2_priors,
                save_pars = save_pars(all = TRUE),
                silent = 2,
                refresh = 0)
```

```
Trying to compile a simple C file
```

```
Running /Library/Frameworks/R.framework/Resources/bin/R CMD SHLIB foo.c
using C compiler: 'Apple clang version 16.0.0 (clang-1600.0.26.4)'
using SDK: 'MacOSX15.1.sdk'
clang -arch arm64 -I"/Library/Frameworks/R.framework/Resources/include" -DNDEBUG   -I"/Libra
In file included from <built-in>:1:
In file included from /Users/spencerau/Library/R/arm64/4.4/library/StanHeaders/include/stan/
In file included from /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/R
In file included from /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/R
/Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library/RcppEigen/include/Eigen/
```

```
  679 | #include <cmath>
      |          ^~~~~~~
1 error generated.
make: *** [foo.o] Error 1
```

# Results

## Question 1

### Frequentist Results for Q1

```
$TOST
[1] "The equivalence test was non-significant, t(314.31) = 0.608, p = 2.72e-01"

$ttest
[1] "The null hypothesis test was non-significant, t(314.31) = -0.477, p = 6.34e-01"

$combined
[1] "NHST: don't reject null significance hypothesis that the effect is equal to zero \nTOST
```

The TOST equivalence test result shows a non-significant p-value of 0.27, indicating there is insufficient evidence to conclude that the two groups are practically equivalent. The t-test result also shows a non-significant p-value of 0.63, meaning there is no evidence to reject the null hypothesis that the two groups have the same mean. The combined result tells us that neither a significant difference nor practical equivalence between Group A and Group B can be established.
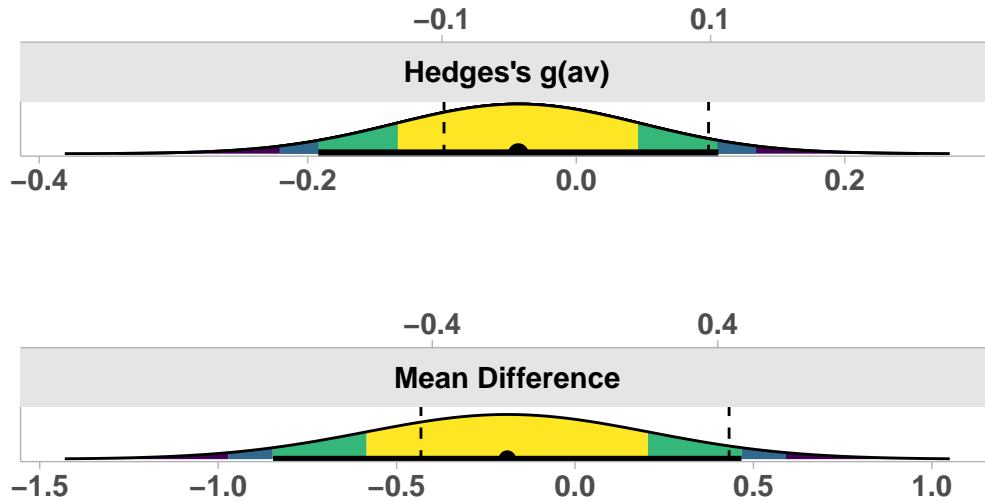
The graph shows the sampling distribution of the mean difference, with the black dot representing the observed mean difference, close to zero, suggesting little difference between the groups. The yellow region, mostly within the equivalence bounds (-0.4, 0.4) indicates some evidence for practical equivalence, but a good portion of the density curve that lies outside the bounds show that equivalence cannot be conclusively established. Overall, the graph supports the conclusion from the t-test and TOST test that we cannot conclude there is a significant difference between Group A and Group B and there is very weak evidence for practical equivalence.

**Bayesian Results for Q1**

```
 q1_fit1_alt q1_fit2_null
    0.125909     0.874091


q1_fit1_alt
  0.1440456


Estimated Bayes factor in favor of q1_fit1_alt over q1_fit2_null: 0.14387
```

The posterior probabilities for the two models are 0.13 for the alternate hypothesis model (where we assume that there is a difference in the mean anxiety score between the two groups) and 0.87 for the null hypothesis model (where we assume that there is no difference in the mean anxiety score between the two groups). Since we assumed that each model has an equal

probability of being true, the Bayes Factor is equal to the posterior odds, which are both 0.14 This means that the alternate hypothesis model is 0.14 times as likely as the null hypothesis model or that the null is 7.14 times as likely as the alternate. This suggests that we have evidence for the null hypothesis model, which is not consistent with the Frequentist analysis of no real conclusion.

We also graph the ROPE (Region of Practical Equivalence) on the fit1 model to determine the percentage of the posterior distribution that falls within the ROPE bounds. We use a ROPE range of -1/10 to 1/10 of the standard deviation of the anxiety scores.

```
# Proportion of samples inside the ROPE [-0.43, 0.43]:

Parameter | inside ROPE
----------------------
Intercept |    59.03 %
groupB    |    71.68 %
```



Region of Practical Equivalence (ROPE)
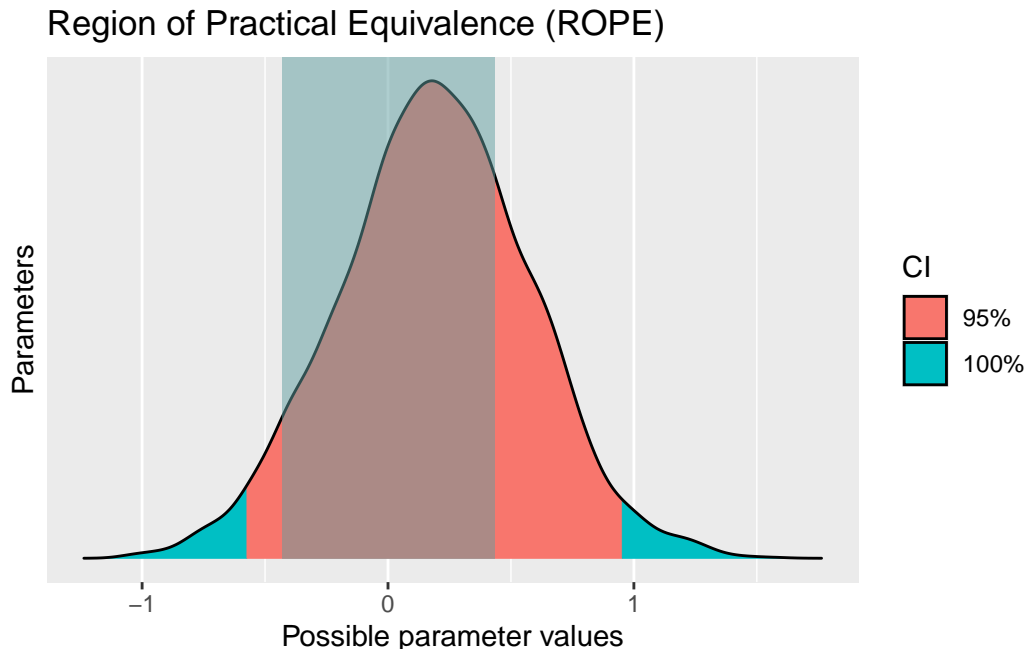
Given that only 57.5% of the posterior confidence interval falls within the ROPE, we cannot conclude that the two groups are practically equivalent given the weak evidence. In addition, only 71.6% of Group B's posterior distribution falls within the ROPE, suggesting that there is some evidence for equivalence but not enough to make a strong conclusion (where we would want something like 90%).

**Answer to Q1**

The Frequentist analysis shows no significant difference (p = 0.63) and no evidence for equivalence (p = 0.27), leaving the results inconclusive. In contrast, the Bayesian analysis provides stronger support for the null hypothesis, with posterior probabilities indicating that no difference is more likely (87%) and moderate evidence for equivalence (57.5%–71.6% in the ROPE). While both approaches suggest the groups may be similar, the Bayesian method quantifies the relative evidence for each model, offering more nuance. The discrepancy arises because Bayesian methods incorporate prior information and provide probabilistic measures of evidence, while Frequentist methods rely on binary decisions based on p-values and hypothesis tests. Ultimately, the Frequentist approach does not have a decisive answer, while the Bayesian approach tells us that the null hypothesis of both groups having the same mean anxiety is more likely to be true.

## Question 2

**Frequentist Results for Q2**

```
Call:
glm(formula = mental_well ~ corgi + age + anxiety, family = gaussian(),
    data = health_data)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.83293    0.39167   4.680 3.71e-06 ***
corgi        0.27334    0.08281   3.301  0.00103 **
age          0.08600    0.01074   8.005 8.54e-15 ***
anxiety     -0.27199    0.02859  -9.514  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 7.409129)

    Null deviance: 5321.8  on 499  degrees of freedom
Residual deviance: 3674.9  on 496  degrees of freedom
AIC: 2426.3

Number of Fisher Scoring iterations: 2
```

The GLM model tells us that there is a statistically significant positive relationship between the number of corgi pets and mental well-being, with a coefficient of 0.27. This means that

for every additional corgi pet per week, mental well-being increases by 0.27. The p value of 0.001 is less than 0.05, indicating that the relationship is statistically significant.

## Effect of Petting Corgis on Mental Well–being (Frequentist)
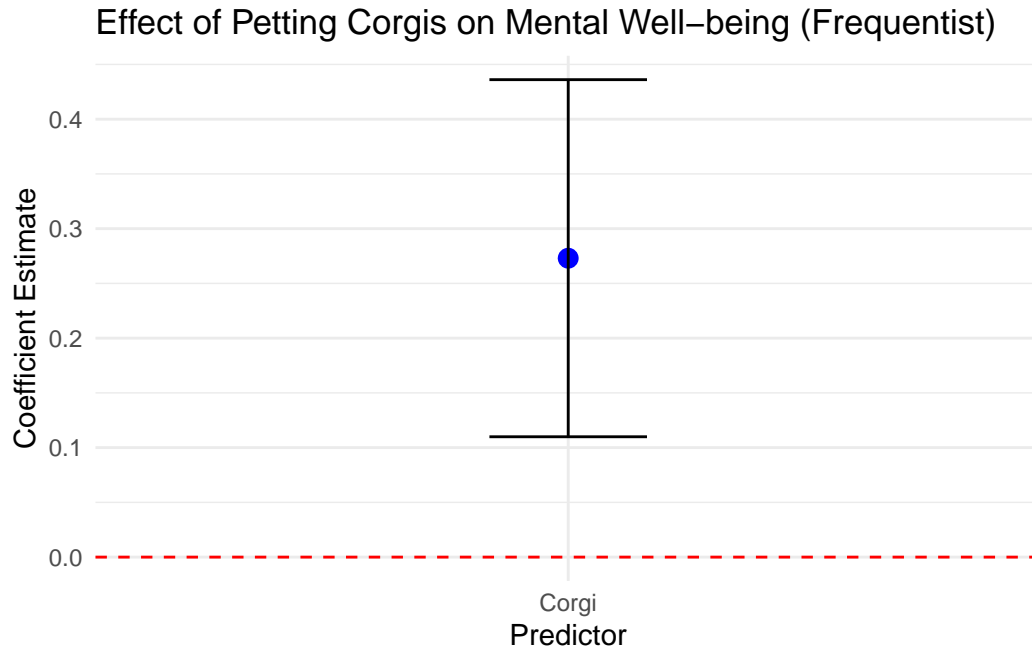


The confidence interval is roughly between 0.1 and 0.45, indicating that 95% of the time, the true effect of petting corgis on mental well-being will fall within this range. The dashed red line represents the null hypothesis that the coefficient is 0, which is clearly rejected by the confidence interval.

**Bayesian Results for Q2**

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: mental_well ~ corgi + age + anxiety
   Data: health_data (Number of observations: 500)
  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 Rhat Bulk_ESS Tail_ESS
Intercept     1.84      0.40     1.06     2.61 1.00     4981     3509
corgi         0.28      0.08     0.11     0.44 1.00     4062     3424
age           0.09      0.01     0.07     0.11 1.00     4295     3234
anxiety      -0.27      0.03    -0.33    -0.22 1.00     4962     3164
```

```
Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     2.73      0.09     2.56     2.91 1.00     5065     2683

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).
```

The Bayesian model estimates that for each additional unit increase in corgi interaction, mental_well increases by approximately 0.27 units on average, with the true effect likely lying between 0.11 and 0.44 given by the lower and upper bounds of the 95% credible interval.

## Posterior Distribution of Corgi Coefficient (Bayesian)
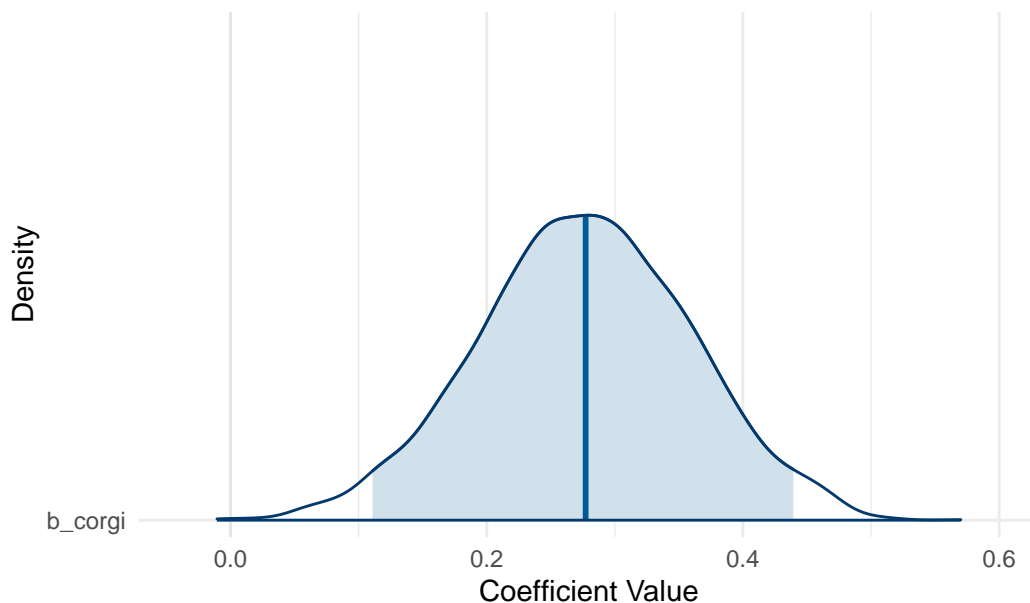


This graph represents the posterior distribution of the corgi coefficient in the Bayesian analysis, showing that the most likely effect of petting corgis on mental well-being is 0.27. The 95% credible interval (0.11 to 0.44) indicates a statistically significant, positive association between petting corgis and mental well-being, as the interval does not include 0.

**Answer to Q2**

The Frequentist and Bayesian analyses both show a statistically significant positive relationship between petting corgis and mental well-being. The Frequentist GLM model estimates that for each additional corgi pet per week, mental well-being increases by 0.27 units, with a 95%

confidence interval of 0.1 to 0.45. The Bayesian model estimates the same effect size of 0.27, with another similar 95% credible interval of 0.11 to 0.44. Both analyses provide strong evidence for a direct, positive causal effect of petting corgis on mental well-being. The results are similar because both methods are estimating the same relationship.

## Discussion

The potential impacts and applications of the answer for Question 1 are that it seems that both Group A and Group B are fairly similar when it comes to mean anxiety measures. This means that neither Andrew nor Betty's methods of leading their respective joy groups have a significant difference in their impact on mean anxiety levels. This could have implications on methods of leadership and techniques geared towards treating anxiety in a group setting. In terms of Question 2, the results suggest that petting corgis has a positive effect on mental well-being. This could have implications for mental health interventions as well as using holistic approaches such as the aforementioned animal therapy towards improving mental well-being.

If I were to redo this assignment, I would probably experiment and play around with a small group of priors for the Bayesian analyses instead of just a single prior