

1. Analysis

For both questions, I conducted both a Bayesian and a Frequentist analysis. For Bayesian models, convergence diagnostics were conducted using `mcmc_trace` plots to ensure reliable estimates from the posterior samples. Additionally, posterior predictive checks (`pp_check`) were applied to validate model fit.

A. Question 1

Question 1 asked me to determine if there is evidence that Groups A and B have the same mean anxiety score.

1.A.i Frequentist Approach

For the Frequentist analysis, I conducted an independent two-sample t-test to assess whether there was any significant difference between anxiety scores for Groups A and B. I used this test because of the continuous nature of the mean anxiety score. This method also has a straightforward interpretation in testing for mean differences between two independent groups, which matches the goal of Question 1. The t-test is a well-suited choice here as it provides a clear hypothesis-testing framework that can be directly interpreted in terms of rejecting or failing to reject the null hypothesis of no mean difference. Additionally, I chose the two-sample t-test over more complex models (such as ANOVA or regression-based approaches) because these would add unnecessary complexity without offering additional insight into a simple two-group comparison. The t-test is computationally efficient, and with a sufficient sample size, it provides an accurate estimate of the mean difference.

There were a few assumptions I made: 1) each observation was independent (supported by the fact that individuals were assigned to groups randomly), 2) the anxiety scores followed a normal distribution, and 3) the variances were homogenous. I will calculate a mean difference in anxiety scores between groups as `mean_diff` with a 95% confidence interval ranging from `ci_lower` to `ci_upper`, meaning that all values within this interval are reasonable guesses of the difference based on the observed data. A conclusion was drawn after looking to see if the interval includes zero, and based on this, I either reject or fail to reject the null hypothesis.

1.A.ii Bayesian Approach

For the Bayesian analysis, I used `brms` to perform a Bayesian analysis with a Gaussian likelihood and a weakly informative prior distribution $N(0, 2.5)$ centered around zero for the group difference (to reflect my moderate uncertainty about the true difference between the two groups) to compare anxiety levels between Groups A and B. I used a Gaussian likelihood because of the continuous nature of the data and assumed symmetric distribution. The prior was chosen because it allows the data to drive the posterior estimation without implying strong initial assumptions. A 95% credible interval was used to estimate plausible ranges for the group difference in anxiety. This allows me to state that there is a 95% chance that the true mean difference lies within this interval.

I did not use other priors (e.g., more informative prior) because they could bias the results if prior knowledge were not accurate or well-established. Instead, the weakly informative prior I chose here respects the data's influence while preventing extreme values. I estimated the posterior mean

difference using $\text{mean}(\text{posterior_samples_q1}\$b_groupB)$. This approach enabled me to generate a probabilistic interpretation of the group differences based on updating beliefs about the parameter after observing the data. If the interval does not contain zero, there is strong evidence for a difference between groups.

B. Question 2

Question 2 required me to find the direct causal effect of petting corgis on mental well-being. To estimate this direct effect, I used generalized linear models for both the Bayesian and Frequentist approaches. Covariates were incorporated based on the DAG provided in the assignment instructions.

1.B.i Frequentist Approach

For the Frequentist approach, I used a generalized linear model coupled with an ordinary least squares estimation because it allowed me to find an unbiased estimate of the direct effect of corgi petting on mental well-being, adjusting for relevant covariates (age, marital status, sex, and anxiety) according to the DAG structure. Specifically, I used the formula $\text{mental_well} \sim \text{corgi} + \text{married} + \text{sex} + \text{anxiety} + \text{age}$. Age was included, as it can influence both mental well-being and the likelihood of engaging in activities like petting corgis. Older individuals might have different mental health baselines and different levels of engagement in pet therapy. Marital status could also be a confounder because it can influence mental well-being through emotional support and social interactions as well as preferences or time availability for activities involving animals. Sex is another confounder since differences in psychological resilience and well-being metrics often vary by sex, and preferences for animal interactions might vary by gender, affecting both the exposure and the outcome. Finally, anxiety was included as it has a direct effect on mental well-being and could also affect how an individual perceives or engages in therapeutic activities like petting corgis. Without adjusting for anxiety, I could inadvertently attribute its effects on mental well-being to corgi petting.

I used a 95% confidence interval from corgi_ci_lower to corgi_ci_upper to estimate the reliability of the effect size. Three assumptions were necessary for my GLM: 1) a linear relationship between average number of corgis pet per week and mental wellbeing (and linear effects from covariates on the outcome), 2) normally distributed errors, and 3) equal variance of errors across values of corgi and covariates (homoscedasticity). Additionally, the Frequentist framework is straightforward to interpret, as each coefficient represents the average change in mental well-being for a one-unit increase in the predictor variable.

1.B.ii Bayesian Approach

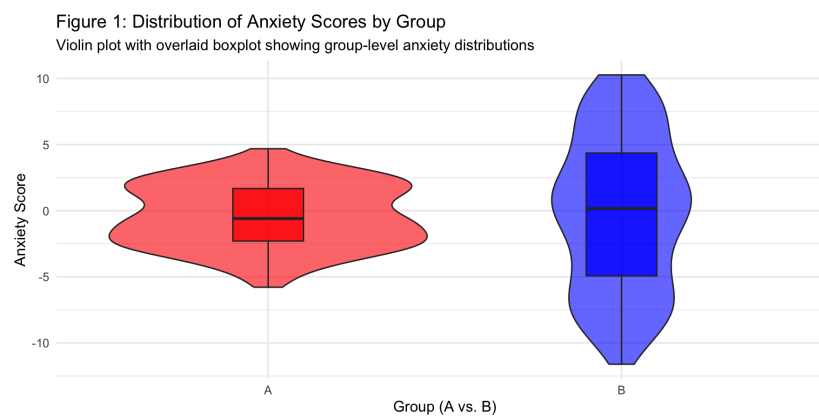
My Bayesian GLM approach, on the other hand, offered flexibility and allowed me to make a probabilistic interpretation of the effect size. I used a Bayesian GLM framework with weakly informative priors due to the plausible variability in mental well-being outcomes. I specified a prior of $N(0, 5)$ for the corgi variable to allow reasonable variation without enforcing strong assumptions. This represented my moderate uncertainty about the expected effect size while ensuring that results were not overly controlled. I also used the same formula as the Frequentist GLM, which allowed me to generate a posterior distribution of the corgi effect on mental well-being. Using this adjustment set in the Bayesian model served the same purpose: to control

for confounding variables that might otherwise bias the estimated direct effect of corgi petting on mental well-being. Each covariate played a specific role in adjusting for various paths that could lead to biased estimates, as explained in my Frequentist approach.

I calculated the Bayesian posterior estimate of the corgi effect as `mean(posterior_samples_q2$b_corgi)`, with a 95% credible interval from `credible_interval_lower_q2` to `credible_interval_upper_q2` (the range in which I believe the true effect lies with 95% probability). The Bayesian analysis also included posterior distribution visualizations, which helped me see the probabilistic range of the corgi effect on mental well-being.

2. Results

To provide a brief visual summary of the data used for my analysis, I plotted the distribution of anxiety scores for employees in Groups A and B (Figure 1). The violin plot shows the density and spread of anxiety scores, with higher density areas indicating more common values. I also added box plots to show the median, interquartile range, and potential outliers for each group. This plot serves to preliminarily explore differences in anxiety levels across groups before statistical testing.



A. Question 1

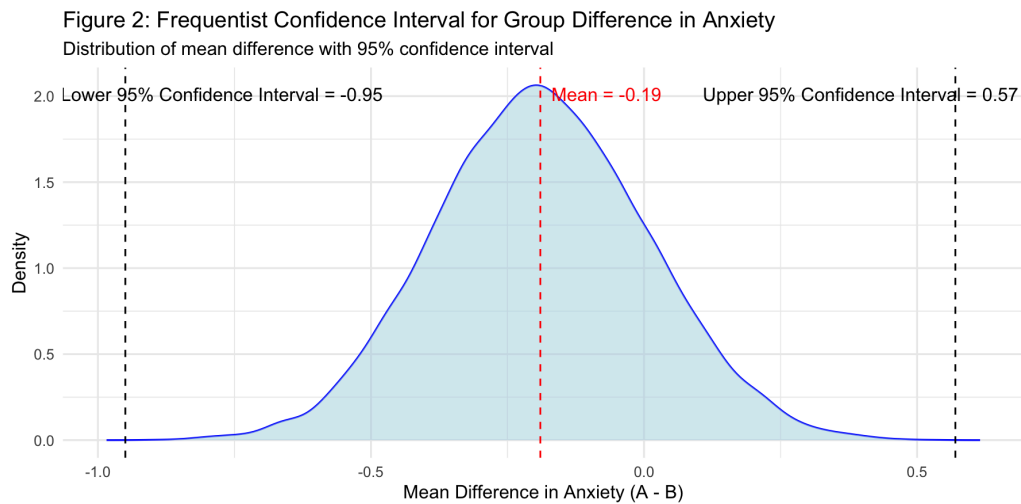
The results from my analyses for Question 1 both indicated that there is no statistically significant difference in mean anxiety between Groups A and B. (There is no evidence that rejects the idea that Group A and Group B have the same mean anxiety score.)

2.A.i Frequentist Approach

In the Frequentist analysis, the two-sample t-test yielded evidence suggesting a slight, non-significant trend toward lower anxiety in Group A compared to Group B. The mean difference was -0.19 with a 95% confidence interval of (-0.95, 0.57) that includes zero, so there is no strong evidence of a true mean difference between groups. Thus, I fail to reject the null hypothesis of no difference; the observed difference may be due to random variation rather than a true effect.

These results can be seen in the density plot in Figure 2. The blue line represents the density of the simulated mean difference distribution. The shaded light blue area marks the 95% confidence interval around the mean difference, from -0.95 to 0.57. The dashed red line indicates the

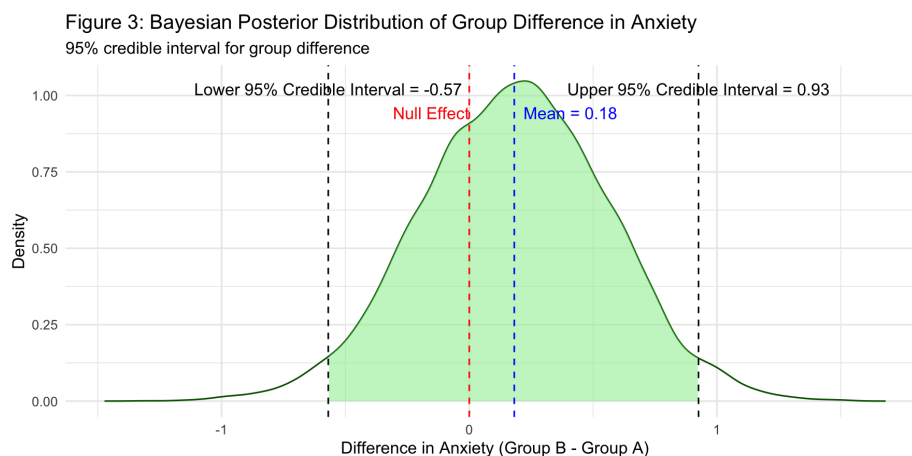
observed mean difference (-0.19), and the dashed black lines mark the confidence interval boundaries.



2.A.ii Bayesian Approach

In the Bayesian analysis, the posterior mean difference (Group A minus Group B) was 0.19, and, while positive, also indicated a slight, non-definitive difference between the groups. The 95% credible interval (-0.57, 0.93) suggests a 95% probability that the true mean difference lies within this interval. Because the interval includes zero, the observed data does not provide strong evidence to suggest that either group leads to a lower anxiety score, consistent with the Frequentist approach.

These results are visualized in Figure 3. The dark green line depicts the density of the posterior samples for the group difference, and the light green shaded region represents the 95% credible interval, extending from -0.57 to 0.93 (marked by the black dashed lines). The dashed blue line marks the posterior mean (0.18), while the dashed red line at zero indicates the null hypothesis of no difference.



After fitting the Bayesian model, I assessed convergence using trace plots for each parameter (Figure 4). The trace plots display sampled values across iterations for the four MCMC chains, allowing us to visually confirm whether the chains have "mixed well." I expect trace plots to

appear like "fuzzy caterpillars" when the chains are converging well, meaning there are no significant upward or downward trends. This pattern suggests that the chains are exploring the posterior distribution efficiently, capturing variability without being trapped in local regions. Here, the overlapping chains without visible trends indicate that my parameter estimates are stable and reliable.

I also conducted a posterior predictive check (PPC) to evaluate model fit (Figure 5). The PPC compares observed data to data simulated from the model's posterior and reflects both distributional variability (uncertainty in parameter estimates) and sampling variability (natural data variation even if we had perfect parameter estimates). In Figure 5, the dark line represents the observed data distribution, while the lighter lines show predictions from the posterior distribution. The close match between these distributions suggests that the model has adequately captured both sources of variability and is therefore a good fit to the data.

Figure 4: Convergence Diagnostics

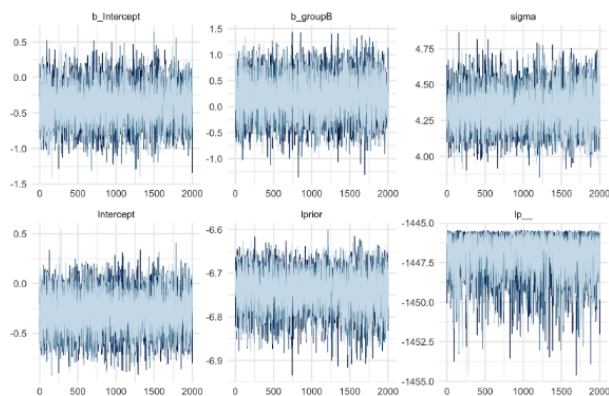
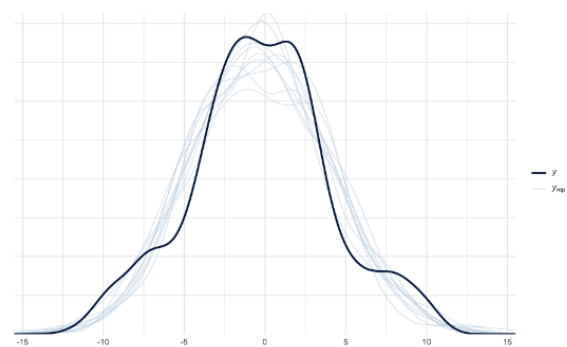


Figure 5: Posterior Predictive Check



2.A.iii Model Comparison

The conclusions of both methods are aligned, as they use the same data (500 individuals split between two groups) and reflect minimal differences in anxiety between the groups. I believe these results align across the two approaches because they both rely heavily on observed data (due to a moderate sample size) and the weakly informative Bayesian prior, which does not strongly influence the posterior distribution. Thus, the data "overpowers" the prior, which led to similar estimates for the mean difference. Additionally, in both methods, I assumed that the data was approximately normally distributed and that variance was consistent across groups.

Because both the Bayesian credible interval and the Frequentist confidence interval include zero, they similarly interpret the plausible range of mean differences, reinforcing the conclusion of no meaningful difference in mean anxiety scores. The considerable overlap between the intervals for both methods indicates robustness in these independent analyses. It suggests that the differences in leadership styles between the two groups do not significantly affect anxiety. Despite the distinct approaches, both methods converge in suggesting no meaningful difference between the groups: the Bayesian credible interval provides a probability-based statement about the difference likely being near zero, while the Frequentist confidence interval frames uncertainty through the frequency of capturing the true parameter in repeated samples.

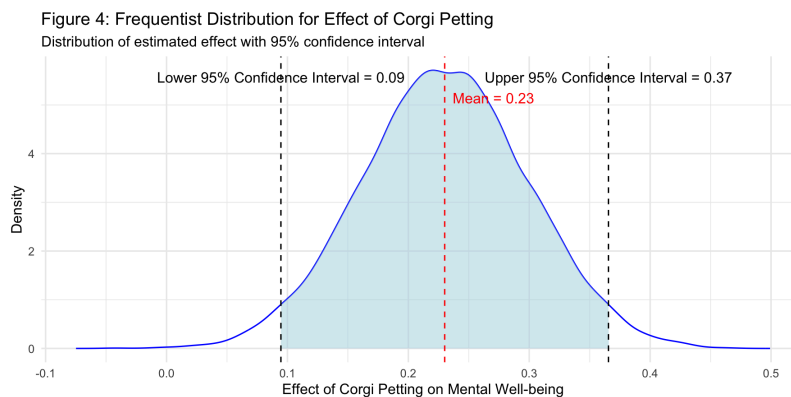
B. Question 2

The results from my analyses for Question 2 both suggested a statistically significant positive direct causal effect of petting corgis on mental well-being.

2.B.i Frequentist Approach

The Frequentist approach resulted in a coefficient of 0.23 (point estimate) for the corgi variable, which suggests that each additional corgi petted per week corresponds to a 0.23-point increase in mental well-being. The 95% confidence interval (0.09, 0.37) excludes zero suggesting that the observed effect of petting corgis on mental well-being is likely positive, with the confidence interval's range reflecting the plausible values for this effect size.

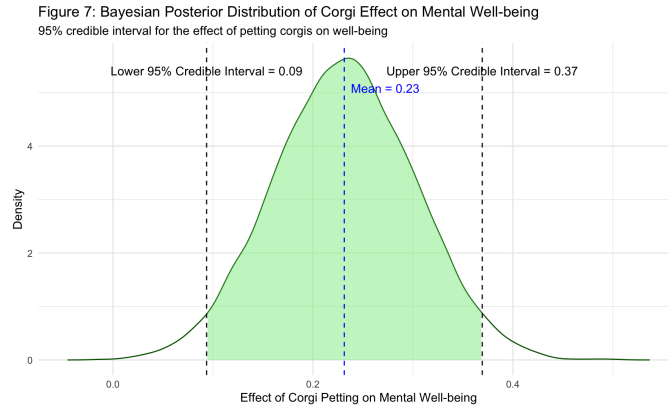
Figure 4 supports these findings. The blue line represents the density of this distribution, and the light blue shaded area highlights the 95% confidence interval, extending from 0.09 to 0.37 (marked by the black dashed lines). The red dashed line at 0.23 marks the estimated mean effect, indicating a positive association between corgi petting and mental well-being.



2.B.ii Bayesian Approach

The Bayesian analysis yielded a posterior mean for the 'corgi' variable of 0.23 (point estimate), with a 95% credible interval of (0.100, 0.365). This interval does not include zero, suggesting that the coefficient for petting corgis has a positive effect on mental well-being. The credible interval allows a probabilistic interpretation that we are 95% confident that the true effect lies between 0.09 and 0.37, supporting a non-zero, positive effect.

The density plot in Figure 5 shows these results. The dark green line represents the posterior density, and the light green shaded area denotes the 95% credible interval from 0.09 to 0.37 (with the black dashed lines showing its upper and lower boundaries). The blue dashed line marks the posterior mean of 0.23.



I also included trace plots for all covariates to check for convergence (Figure 8). Again, the trace plots should ideally resemble "fuzzy caterpillars," indicating stable sampling around the parameter means. These trace plots suggest good mixing and dependable posterior estimates because the chains overlap and oscillate without trends. This indicates that the chains have thoroughly explored the posterior space and provided reliable parameter inferences.

The posterior predictive check (Figure 9) further verifies my model's fit by comparing observed data with predictions generated from the posterior. This PPC also captures two types of variability (distributional and sampling, as in Figure 5). The dark line in Figure 9 shows the observed distribution, while the light lines represent predictive distributions. Their alignment indicates a strong model fit because it captures both the uncertainty in parameter estimates and the inherent randomness of data.

Figure 8: Convergence Diagnostics

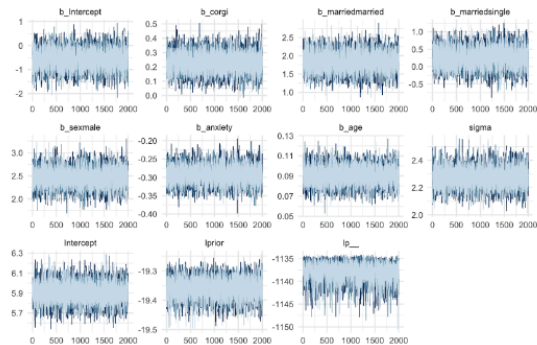
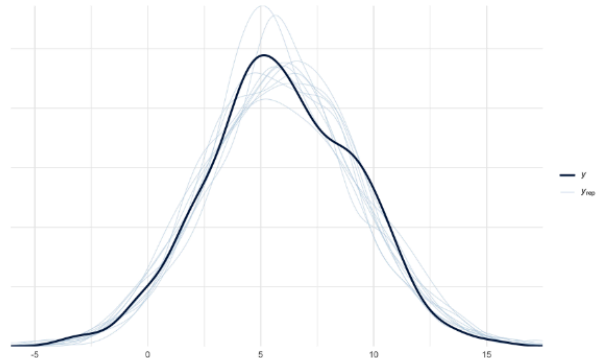


Figure 9: Posterior Predictive Check



The Frequentist confidence interval and the Bayesian credible interval have significant overlap, and both contain zero, which adds credibility to my claim of a positive effect. The Bayesian credible interval directly supports the probability of the effect being positive, which aligns with my conclusion from the Frequentist approach.

2.B.iii Model Comparison

For both questions, my Frequentist confidence intervals and Bayesian credible intervals provide consistent conclusions, which reinforces the validity of my findings. I believe these findings were consistent because, similar to Question 1 both frameworks used observed data for the

estimated effect of corgi petting on mental well-being. Since the sample size and signal in the data were substantial enough, both methods produced similar central estimates for the corgi effect. This meant that while the Bayesian prior moderated the effect slightly, it did not significantly shift the posterior estimate away from the Frequentist estimate, which is purely data-driven.

The Bayesian prior for the corgi effect was set as $N(0,5)$, which is weakly informative and centered around zero. This prior allows for a wide range of possible effect sizes without enforcing a strong bias toward any particular outcome. Consequently, the observed data had enough weight to dominate the prior, resulting in a posterior distribution that closely mirrored the Frequentist estimate. If a stronger prior were used, the results might have diverged more.

Both analyses also used the same model structure with covariates (corgi, married, sex, anxiety, and age) to control for other potential influences on mental well-being. This consistency in covariate adjustment allowed both approaches to focus on isolating the corgi effect which is another reason why I believe the two methods resulted in a similar central estimate despite different statistical philosophies.

The Frequentist GLM assumed normally distributed residuals, and the Bayesian model used a Gaussian likelihood, which assumed a similar distribution for the observed data. This alignment in distributional assumptions likely also contributed to similar estimates for the corgi effect in both methods.

Both methods suggest no significant difference in anxiety scores between groups in Question 1 while identifying a positive association between petting corgis and mental well-being in Question 2. The main distinction between the two approaches lies in their interpretation of uncertainty: the Bayesian credible interval provides a probability-based statement about the difference likely being near-zero, while the Frequentist confidence interval frames uncertainty through the frequency of capturing the true parameter in repeated samples.

3. Discussion

A. Potential Impacts/Applications

While there was no statistically significant difference in anxiety outcomes between the two groups, Andrew and Betty could use this information to assess other potential impacts on anxiety levels and perhaps assist in informing their leadership training. Additionally, Company X may not need to modify the joy group assignments based only on anxiety outcomes. This outcome could encourage the company to explore other factors like group cohesion that could potentially impact anxiety levels between the two groups.

The conclusion that petting corgis positively affects mental well-being (adjusted for other variables) could encourage the company to provide regular pet therapy sessions as a part of their wellness program. Additionally, the company might consider exploring the personalization of its programs to account for employee preferences regarding animal interactions. Furthermore, the results of question 2 present an opportunity for the company to assess whether sustained exposure to animal interactions could maintain or even increase mental well-being over an extended timeframe via a longitudinal study. While petting corgis is shown to positively affect mental well-being, other covariates from the experiment's DAG (such as marital status or age) may also have noteworthy effects on mental well-being.

B. If I Redid the Assignment

If I were to redo the assignment, I might run an actual experiment to collect empirical data, as synthetic data may not fully capture the characteristics of the real population. Thus, collecting and using real-world data could help improve the plausibility of my claims and would, therefore, be more helpful to stakeholders (e.g., the company) and potentially generalizable to a measured extent. Additionally, if I redid the assignment, I could test different priors in my Bayesian analysis to verify the robustness of the results under various assumptions. For instance, using a narrower normal prior with a lower standard deviation could reduce the credible interval's width and reflect greater confidence in the prior knowledge. Since this cross-sectional design only captures a single time point, a longitudinal study might better support causal inference, especially for analyzing the long-term effects of interventions like pet therapy on mental well-being. Finally, while the DAG for this assignment shows various relationships between factors that impact mental well-being, if I redid the assignment, I could expand this DAG to explore additional variables and therefore potential mediators, confounders, and colliders.