Research Question 2: Count Model Regression

Objective of Analysis

This study investigates how psychological and behavioural factors influenced daily physical exercise frequency during the COVID-19 pandemic. The pandemic context, marked by isolation and restrictions, likely amplified the role of these factors, shaping exercise behaviours and coping mechanisms (**?**). Findings offer insights into how dietary habits and loneliness impact exercise frequency, informing targeted public health strategies to promote healthier behaviours and mitigate the negative effects of prolonged restrictions (**?**).

The data analysed are published in **?** and can be found in the CSV file *'loneliness_original.csv'* inside the data sub-directory. The dataset focuses on individuals' behavioural and psychological attributes during the COVID-19 pandemic in Spain and was collected between March and May 2020. The data is comprised of 3421 participants, with the count outcome variable, **Physical Exercise**, which represents the frequency of an individual partaking in daily exercise ($mean$ = 2.27 times per day; SD = 0.92; range= 1 to 4). Key psychological factors include **Loneliness**, a continuous variable ($median$ = 19;$mean$ = 19.43; SD = 5.75; range = 11 to 43) reflects the degree of perceived social isolation among participants, with higher scores indicating greater levels of loneliness. Behavioural variables include **Healthy Diet**, a categorical variable of how healthy an individual's diet is out of 1 to 4, with 4 being the healthiest (least sugar) and 1 being the least healthy (most sugar) ($mean$ = 2.81; SD = 0.8; range = 1 to 4). Demographics variables comprise: **Age**, a continuous variable, ranging from 17 and 84 years ($median$ = 35; $mean$ = 36.88; SD = 15.21). As well as, **Gender**, a categorical variable with the majority of participants identifying as female (n = 2455; 71.76%) compared to male (n = 966; 28.24%).

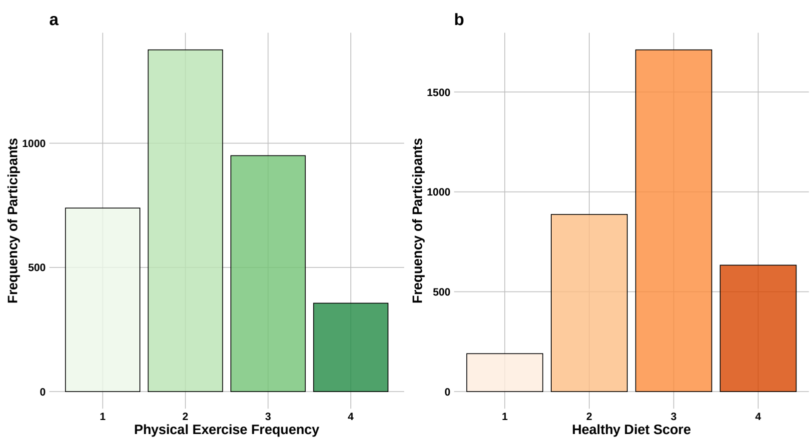

Figure 1: The Distribution of Physical Exercise Frequency and Healthy Diet Scores among Participants.

Figure **??**:a illustrates most participants exercise twice (40.22%; n = 1376) per day, followed by three times (27.77%; n = 950) per day, once (21.6%; n = 739) per day, and four times (10.41%; n = 356) per day. The healthy diet bar plot (b) highlights most participants rated their diet as 3 (50.01%; n = 1711), followed by a rating of 2 (25.93%; n = 1376), a rating of 4 (10.41%; n = 356), and a rating of 1 (21.6%; n = 739).

Statistical model

To model the count outcome of physical exercise, different statistical approaches were considered. The Poisson regression model, suitable for count data with equal mean and variance, was initially tested but rejected due to evidence of underdispersion ($\phi = 0.315$, where $\phi < 1$), which violates the Poisson assumption and risks overestimating standard errors. The Negative Binomial model, designed for overdispersed data, and the Zero-Inflated Negative Binomial model were also unsuitable, as the data exhibited underdispersion and contained no zero counts. The Quasi-Poisson model was selected as the most appropriate approach. Unlike the Negative Binomial model, the Quasi-Poisson model introduces a dispersion parameter (ϕ) that adjusts the variance independently of the mean ($\text{Var}(Y) = \phi \cdot \mu$), providing flexibility to handle underdispersion. This model retains the log-link function, preserving coefficient interpretability on the log scale, where predictors represent multiplicative effects on the outcome. To capture nonlinear relationships, quadratic terms were incorporated, and interaction terms (e.g., healthy diet $\times$ loneliness) were tested to explore potential dependencies between predictors. For example, the interaction between diet quality and loneliness examines whether individuals with better diets are more likely to exercise, depending on their level of social connectedness. The final Quasi-Poisson regression model was specified as:

$$\log(\mu_i) = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}$$

where μ_i is the expected physical exercise frequency for individual i, β_0 is the intercept, β_k are the coefficients for predictors and x_{ki} is the value of predictor k for individual i. A one-unit increase in x_{ki} leads to a proportional change in μ_i by a factor of $\exp(\beta_k)$, holding other predictors constant. The model was implemented in R using the glm() function:

```
quasi_poisson_model <- glm(physical_exercise ~ healthy_diet * loneliness +  gender
                           + age + I(age^2), family = quasipoisson, data = covid_df)
```

Influential points identified using Cook's Distance ($4/n$) were removed to improve robustness and variance inflation factor analysis revealed no significant multicollinearity concerns among predictors. Residual diagnostics indicated a good fit, with no major patterns observed in residual plots.

Comparison of Models

```
poisson_model <- glm(physical_exercise ~ healthy_diet * loneliness + gender
                           + age + I(age^2), family = poisson, data = covid_df)

negative_binomial_model <- glm.nb(physical_exercise ~ healthy_diet * loneliness
                                  + gender + age + I(age^2), data = covid_df)

anova(poisson_model, negative_binomial_model, test = "Chisq")
```

Table 1: Model Comparison Results.

Model	AIC	BIC	Deviance	McFadden_R2
Poisson	9667.45	9710.14	920.43	0.03
Negative Binomial	9669.49	9718.28	920.41	0.03

Model comparison (Table **??**) reveals that the Poisson model has the lowest AIC (9667.45), but underdispersion in the data makes Quasi-Poisson model a more reliable choice, as it adjusts for underdispersion while maintaining the same deviance (920.43) as the Poisson model. The Negative Binomial model has a slightly lower deviance (920.41), indicating a marginally better fit, but it is less suitable due to its higher AIC (9669.49) and overdispersion handling, which is unnecessary for this dataset. Both Poisson and Negative Binomial models have McFadden's R^2 of 0.03, indicating modest model fit improvement. Despite the low R^2, the models still provide valuable insights into the relationships between predictors and physical exercise. A Drop1 test indicated that removing the quadratic term for age significantly increases deviance and AIC, underscoring its contribution to model fit. The interaction between healthy diet and loneliness also plays a role, suggesting that diet's effect on physical exercise may depend on loneliness. The ANOVA comparison between the Poisson and Negative Binomial models indicates that, despite the Poisson model's slightly better AIC, the Negative Binomial model offers a better fit with lower deviance. Therefore, Quasi-Poisson regression, which incorporates a quadratic term, is chosen due to the model's ability to account for underdispersion while capturing nonlinear relationships (**?**).

Model Evaluation and Interpretation

Table 2: Coefficients Table for Quassi-Poisson Model.

	Estimate (Beta)	Std. Error	t-value	p-value	exp(Beta)	95% CI (exp(Beta))
(Intercept)	0.58	0.10	5.83	< .001***	1.78	[1.47, 2.16]
Healthy Diet	0.24	0.03	8.27	< .001***	1.27	[1.2, 1.34]
Loneliness	-0.01	0.00	-1.40	0.161	0.99	[0.99, 1]
Gender: Female	0.01	0.01	0.40	0.688	1.01	[0.98, 1.03]
Age	-0.02	0.00	-6.57	< .001***	0.98	[0.98, 0.99]
Age (Quadratic)	0.00	0.00	5.64	< .001***	1.00	[1, 1]
Healthy Diet X Loneliness	0.00	0.00	-0.16	0.874	1.00	[1, 1]

Note. Significance levels: $***\ p < .001$, $**\ p < .01$, $*\ p < .05$.

The Quasi-Poisson regression model highlights the key predictors influencing physical exercise frequency, with strong evidence for some predictors and inconclusive findings for others (Table **??**). The model's intercept ($\beta_0 = 0.58$, $p < 0.001$) represents the log-expected frequency of physical exercise when all predictors are at their mean or reference values. After exponentiation, $\exp(\beta_0) = 1.78$, indicates that individuals engage in physical exercise approximately 1.78 times per day on average.

Healthy Diet emerges as the most significant predictor, showing a positive association with physical activity ($\beta = 0.24$, $p < 0.001$). A one-unit improvement in diet quality corresponds to a 26.71% increase in exercise frequency ($\exp(\beta) = 1.27$, 95% CI: [1.2, 1.34]). These findings highlight the critical role of dietary quality in promoting exercise and its potential as a public health intervention target. **Age** exhibits a significant nonlinear relationship with physical exercise. The linear term ($\beta = -0.02$, $p < 0.001$) indicates a decline in exercise frequency with increasing age. However, the quadratic term ($\beta = 0$, p <0.001) reveals that this decline stabilizes or reverses in older populations. These results highlight the need for age-specific interventions, particularly targeting younger populations to address declining exercise habits. **Loneliness** ($\beta = -0.01$, $p = 0.16$) was not a significant predictor, as its confidence intervals (95% CI: [0.99, 1]) overlap 1. This suggests a negligible effect on physical activity, possibly due to variability in how loneliness influences behaviour—some individuals may use exercise to cope, while others may become more sedentary. The small effect size may also contribute to non-significance. Similarly, **Gender (Female)** ($\beta = 0.01$, $p = 0.69$) showed no substantial differences between males and females (95% CI: [0.98, 1.03]). This result likely reflects balanced exercise behaviours during the pandemic. The interaction between **Healthy Diet and Loneliness** ($\beta = 0$, $p = 0.87$) was also non-significant (95% CI: [1, 1]), indicating that the relationship between diet quality and physical activity does not depend on loneliness levels, suggesting the effect of diet quality on physical activity is additive.

The standard errors for significant predictors, such as healthy diet ($SE = 0.03$) were small, indicating precise estimates. Likewise, the small standard error for loneliness ($SE = 0$) reflects high precision, with its non-significance likely attributable to a small effect size or limited variability rather than estimation error. The model's explanatory power is supported by a reduction in deviance from 1183.7 (null deviance) on 3290 degrees of freedom to 920.4 (residual deviance) on 3284 degrees of freedom. This reduction indicates that the included predictors substantially improve the model's fit.

Model predictions

```r
predict(quasi_poisson_model, newdata = prediction_data, type = "link")
```

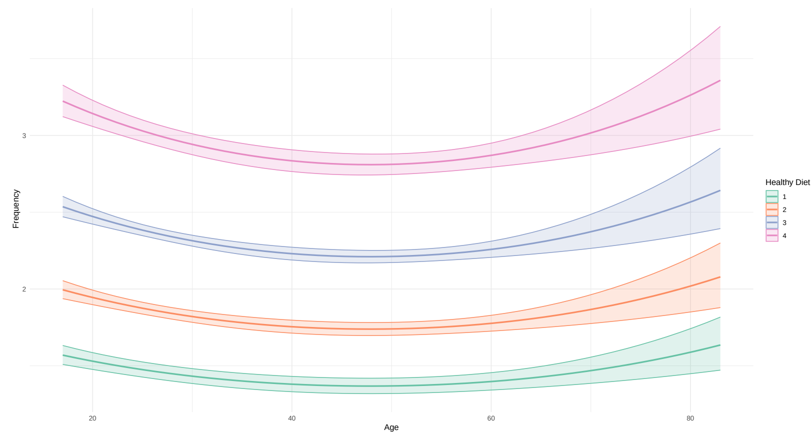

Figure 2: Predicted Exercise Frequency by Age and Healthy Diet.

Figure **??** illustrates how predicted exercise frequency varies across the levels of a healthy diet, higher levels are associated with increased predicted exercise frequency. This suggests a positive relationship between diet quality and exercise frequency, indicating that individuals with healthier diets tend to have higher levels of physical activity. This aligns with health behaviour theories, which suggest that structured routines and positive habits, such as maintaining a healthy diet, can mitigate the impact of stress and support higher levels of physical activity, particularly during crises like the COVID-19 pandemic. The line plot provides further insights by showing how predicted exercise frequency changes with age, stratified by levels of healthy diet. Exercise frequency decreases or plateaus at certain ages for some diet levels while increasing for others. This highlights potential interactions between age and diet quality, suggesting that the relationship between diet and exercise frequency may vary across different age groups.

Conclusion

This analysis investigated how psychological and behavioural factors influenced physical exercise frequency during the COVID-19 pandemic using a Quasi-Poisson regression model to account for underdispersion. Healthy diet emerged as the most significant predictor, with a healthier diet strongly associated with increased exercise frequency. Age exhibited a nonlinear relationship with exercise, highlighting a decline in activity levels with age, which stabilized or reversed in older populations. Loneliness and gender did not significantly influence exercise patterns, suggesting that individual coping strategies may moderate its effects. However, several limitations must be noted. The bounded range of the exercise variable (1- 4) limits variability, potentially underestimating higher activity levels. The cross-sectional nature of the data prevents causal inferences about the relationship between diet, loneliness, and exercise. Potential response bias in self-reported exercise and diet data could distort associations. Nonetheless, these findings emphasize the importance of promoting healthy diets as a strategy to increase physical activity, especially during periods of isolation. Tailored interventions targeting younger populations to counteract declining exercise habits, alongside age-specific approaches for older adults, could enhance physical and mental health. Future research should explore longitudinal data and objective measures of behaviour to validate these findings and refine public health strategies, particularly by examining interactions between diet quality, loneliness, and age. These insights can inform policies to foster resilience and healthy behaviours during prolonged isolation.