My title*

My subtitle if needed

Mary Cheng

March 27, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

2 Data

Some of our data is of penguins (**?@fig-bills**), from Horst, Hill, and Gorman (2020).

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

3 Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix A.

*Code and data are available at: https://github.com/marycx/us_poverty_analysis_2019.git

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) &(1)\\
\text{logit}(\pi_i) &= \alpha + \beta_1 \times \text{mortgage}_i + \beta_2 \times \text{income}_i + \beta_3 \times \text{age}_i &(2)\\
\alpha &\sim \text{Normal}(0, 2.5) &(3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) &(4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) &(5)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in Table 1.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Table 1: Explanatory model Poverty Prediction (n = 1000)

	In poverty
(Intercept)	−7.573
	(1.768)
mortgageOwner without mortgage	−0.028
	(0.441)
mortgageRenter	1.105
	(0.411)
income10k-50k	5.957
	(1.757)
income150k-200k	−2.265
	(4.196)
income200k-250k	−2.798
	(4.922)
income50k-100k	1.819
	(1.825)
incomeabove 250k	−2.677
	(4.703)
incomebelow 10k	21.212
	(9.366)
age35-44	−0.023
	(0.437)
age45-54	0.230
	(0.519)
age55-64	0.203
	(0.504)
ageabove 65	0.644
	(0.454)
Num.Obs.	1000
R2	0.433
Log.Lik.	−159.015
ELPD	−168.0
ELPD s.e.	13.1
LOOIC	336.1
LOOIC s.e.	26.3
WAIC	336.0
RMSE	0.23

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Model details

A.1 Posterior predictive check

In Figure 1, we implement a posterior distribution. This compares the poverty status of people in reality with the prediction results from the posterior distribution from our logistic regression model. It can be seen that the posterior distribution fits well with the actual data. It suggests that the posterior is able to generate simulated data that closely resembles the actual data (**bayes_post?**), because the model accurately captures the observed data patterns. This is good because it indicates that our logistic regression model is a good representation of the actual poverty status in the 2019 poverty data from United States Census Bureau.

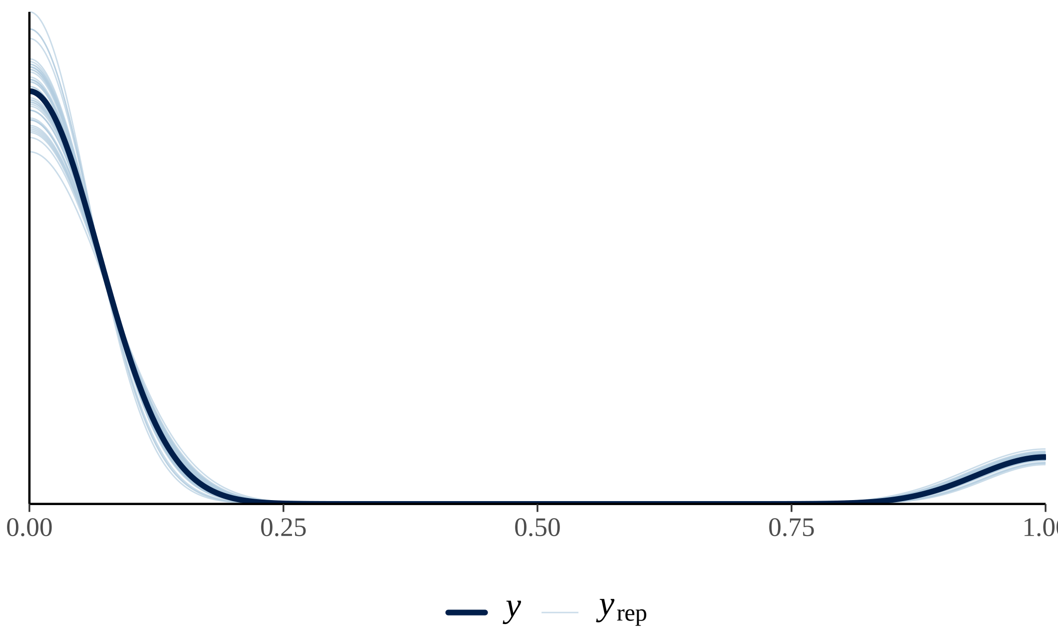


Figure 1: Posterior distribution for logistic regression model

Figure 2 compares the posterior with the prior. This compares the prior distribution of parameters with the posterior distribution of parameters in our logistic regression model. We can see that half of the model parameters do not change after data are taken into account, while some parameters shift slightly. This shows that the observed data partially matches with our initial belief and expectation about the poverty status of people in the United States in 2019. We can see that for people with income level at “below 10k”, the posterior distribution greatly shifts from its prior after we input observed data. This is likely suggesting that the observed data for “below 10k” strongly contradicts our initial belief. This suggests that the majority of people in US who earns less than a total amount of 10k per household annually is in poverty in 2019.

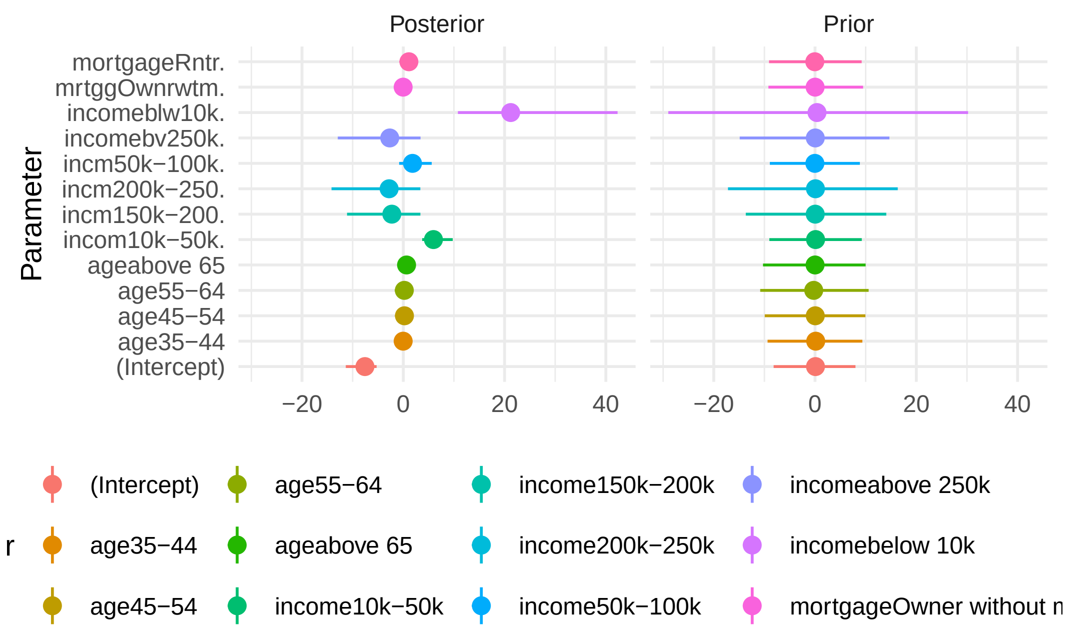
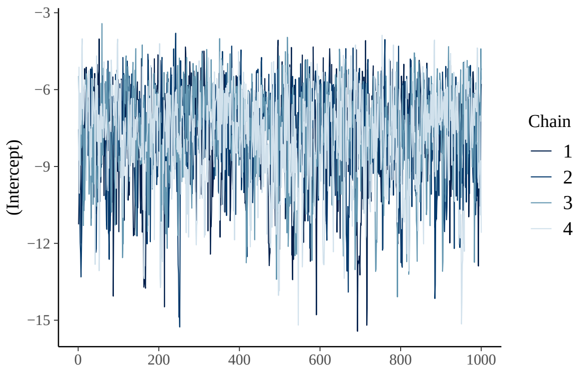


Figure 2: Comparing the posterior with the prior

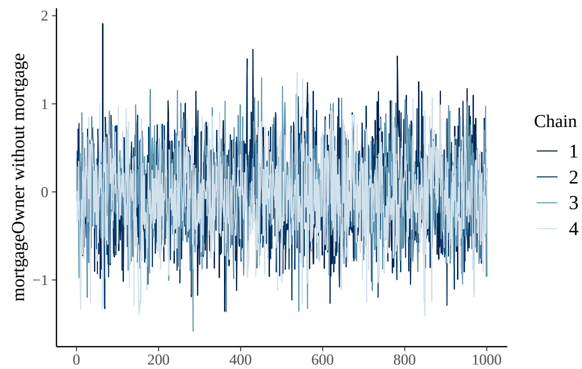
A.2 Markov chain Monte Carlo Convergence Check

Figure 3, Figure 4, Figure 5 are the trace plots of the model. It tells us if there is existence of signs that the our model runs into issues. We observe lines in all the trace plots are horizontal and oscillating, and have overlaps between the chains. This suggests that there is nothing strange in this trace plot.

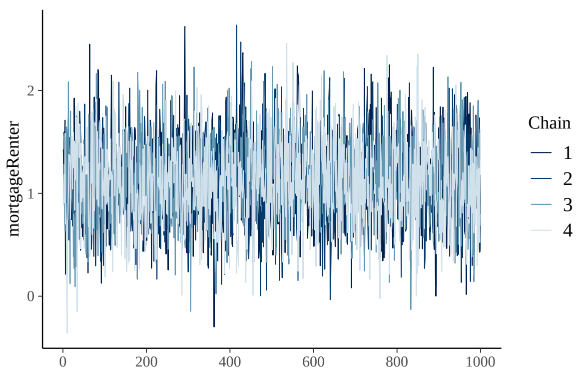
Figure 6 is the Rhat plot of the model. It compares the variability within each chain to the variability between chains in MCMC. We can observe that our Rhat plot are all close to 1, and no more than 1.05. This is a good sign because it suggests that the MCMC algorithm has reached convergence for our model.



(a) Trace plot of Intercept

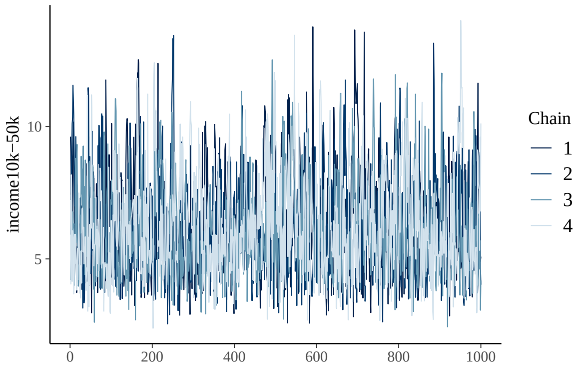


(b) Trace plot of owner without mortgage

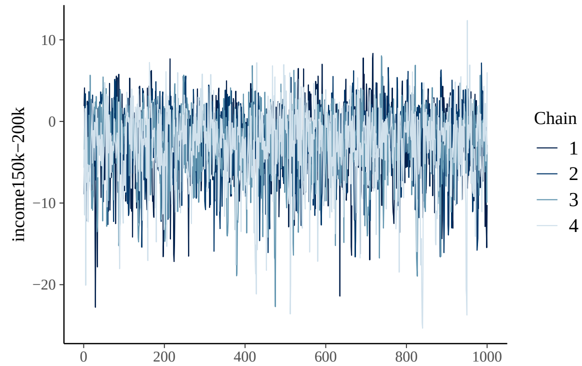


(c) Trace plot of renter

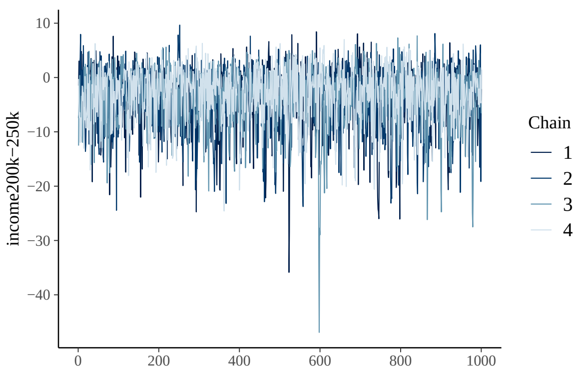
Figure 3: Trace plot of intercept and mortgage



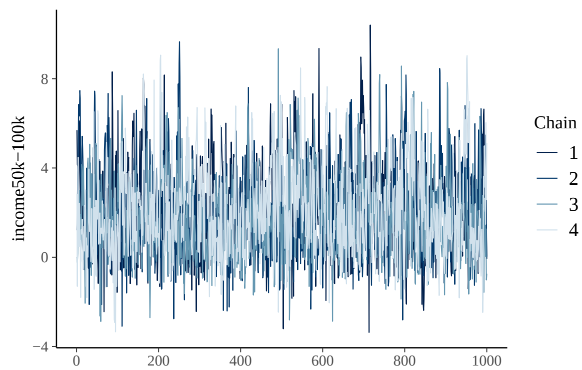
(a) Trace plot of income 10k-50k



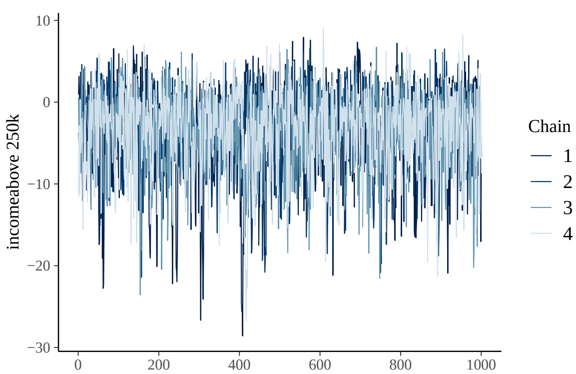
(b) Trace plot of income 150k-200k



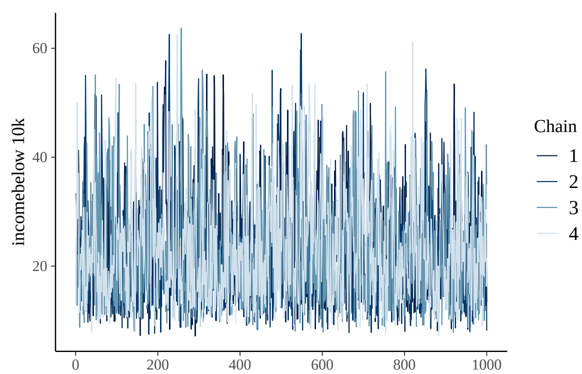
(c) Trace plot of income 200k-250k



(d) Trace plot of income 50k-100k

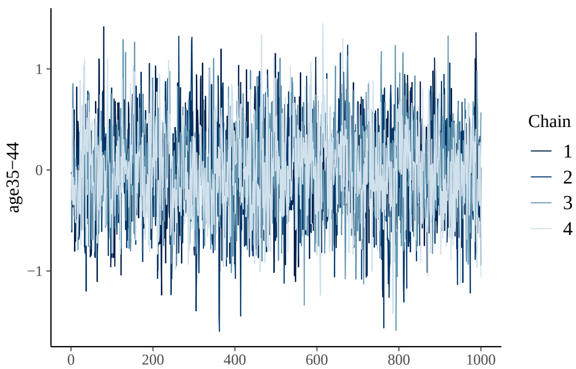


(e) Trace plot of income above 250k

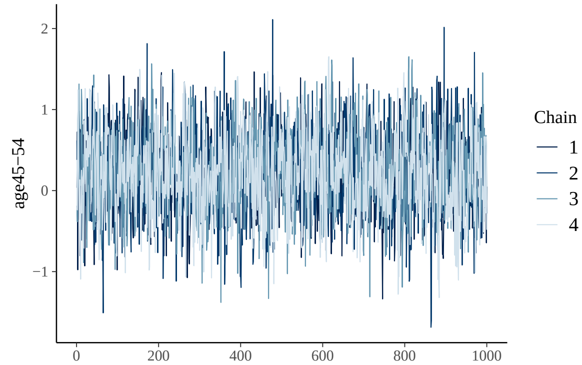


(f) Trace plot of income below 10k

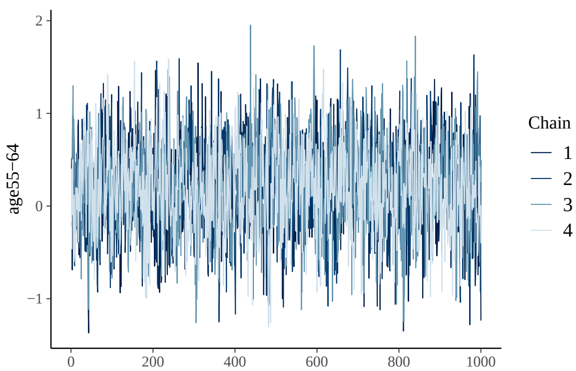
Figure 4: Trace plot of income



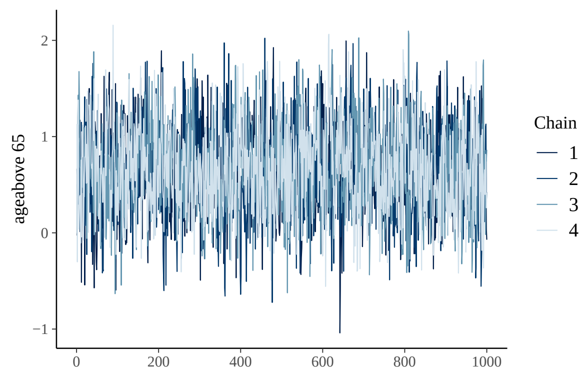
(a) Trace plot of age 35-44



(b) Trace plot of age 45-54



(c) Trace plot of age 55-64



(d) Trace plot of age above 65

Figure 5: Trace plot of employ

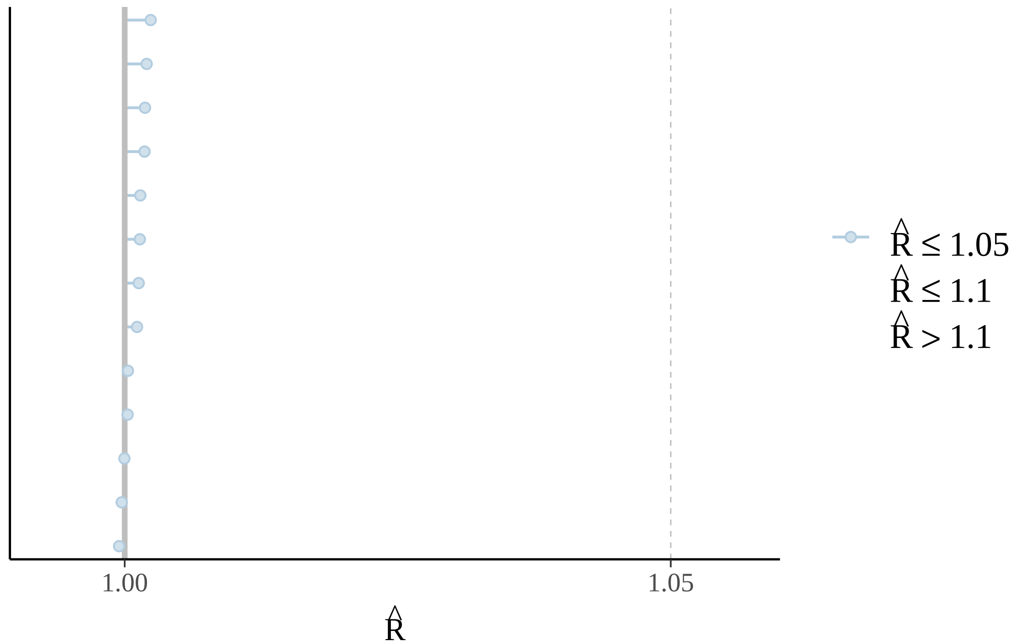


Figure 6: Rhat plot

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. https://doi.org/10.5281/zenodo.3960218.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.